

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
26 July 2001 (26.07.2001)

PCT

(10) International Publication Number
WO 01/53478 A2

(51) International Patent Classification⁷: C12N 15/00

(74) Agents: KOWALSKI, Thomas, J. et al.; Frommer
Lawrence & Haug LLP, 745 Fifth Avenue, New York, NY
10151 (US).

(21) International Application Number: PCT/US01/02051

(22) International Filing Date: 22 January 2001 (22.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
001578.4 24 January 2000 (24.01.2000) GB
1 7 December 2000 (07.12.2000) US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(71) Applicants (*for all designated States except US*): GEN-
DAQ LIMITED [GB/GB]; 103 Burtonhole Lane, London
NW7 1AD (GB). ROCKEFELLER UNIVERSITY
[US/US]; 1230 York Avenue, Box 62, New York, NY
10021 (US).

Published:

— without international search report and to be republished
upon receipt of that report

(71) Applicants and

(72) Inventors: CHOO, Yen [GR/GB]; MRC Laboratory of
Molecular Biology, Medical Research Council Centre,
Hills Road, Cambridge CB2 2QH (GB). ULLMAN,
Christopher, Graeme [GB/GB]; Gateway Mews, Shack-
lewell Lane, London E8 2DF (GB). CHUA, Nam [US/US];
Rockefeller University, 1230 York Avenue, Box 62, New
York, NY 10021 (US). SANCHEZ, Juan, Pablo [US/US];
Rockefeller University, 1230 York Avenue, Box 62, New
York, NY 10021 (US).

(48) Date of publication of this corrected version:

8 November 2001

(15) Information about Correction:

see PCT Gazette No. 45/2001 of 8 November 2001, Section
II

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

(54) Title: REGULATED GENE EXPRESSION IN PLANTS, PROMOTER SYSTEMS FOR USE THEREIN AND METHODS
OF USE THEREOF

(57) Abstract: A method is provided of regulating transcription in a plant cell from a DNA sequence comprising a target DNA
operably linked to a coding sequence, which method comprises introducing an engineered zinc finger polypeptide in said plant cell
which polypeptide binds to the target DNA and modulates transcription of the coding sequence.

WO 01/53478 A2

TITLE**REGULATED GENE EXPRESSION IN PLANTS, PROMOTER SYSTEMS
FOR USE THEREIN AND METHODS OF USE THEREOF****5 REFERENCE TO RELATED APPLICATIONS/INCORPORATION BY
REFERENCE**

This application is a continuation-in-part of US patent application serial number 09/732,348 filed December 7, 2000 which is a continuation-in-part of PCT application no. PCT/GB00/02071 entitled "GENE SWITCHES" filed 30 May 2000 designating the US and claiming priority from GB applications 9912635.1 filed 18 May 1999 and 001578.4 filed 24 January 2000. Further mentioned and incorporated by reference herein are PCT/GB99/03730, filed November 9, 1999, published as WO00/27878A1 on May 18, 2000 entitled "Screening System For Zinc Finger Polypeptides For A Desired Binding Ability" and claiming priority from GB application 9824544.2, filed November 9, 1998, and designating the US; PCT/GB99/03730 which is a continuation-in-part of US patent application Serial No. 09/139,672, filed August 25, 1998 (now US Patent No. 6,013,453), which is a continuation of US patent application Serial No. 08/793,408 (now US Patent No. 6,007,988), filed as PCT application no. PCT/GB95/01949 on August 17, 1995, designating the U.S. and, published as WO96/06166 on February 29, 1996 entitled "Improvements in or Relating to Binding Proteins for Recognition of DNA"; PCT/GB95/01949 claims the benefit of priority from GB application 9514698.1, filed July 19, 1995, GB application 9422534.9, filed November 8, 1994 and GB application no. 9416880.4, filed August 20, 1994. Mention is also made of: USSN 08/422,107; WO96/32475; WO99/47656A2, published September 23, 1999 entitled "Nucleic Acid Binding Proteins"; WO98/53060A1, published November 26, 1998 entitled "Nucleic Acid Binding Proteins"; WO98/53059A1 published November 26, 1998 entitled "Nucleic Acid Binding Proteins"; WO98/53058A1 published November 26, 1998 entitled "Nucleic Acid Binding Proteins"); WO98/53057A1 published November 26, 1998 ("Nucleic Acid Binding Polypeptide Library"; US Patent Nos. 6,013,453 and 6,007,988; Fiehn et al. (2000) Nature Biotechnol. 18:1157-1161; Richter et al. (2000) Nature Biotechnol. 18:1167-1171; and,

generally, Nature Biotechnol. Vol. 18(11) together with all documents cited or referenced therein. Each of the foregoing applications and patents, and each document cited or referenced in each of the foregoing applications and patents, including during the prosecution of each of the foregoing applications and patents
5 ("application cited documents") and any manufacturer's instructions or catalogues for any products cited or mentioned in each of the foregoing applications and patents and in any of the application cited documents, are hereby incorporated herein by reference. Furthermore, all documents cited in this text, and all documents cited or referenced in documents cited in this text, and any manufacturer's instructions or
10 catalogues for any products cited or mentioned in this text, are hereby incorporated herein by reference.

TECHNICAL FIELD

This invention relates to the regulation of gene expression in plants using engineered zinc finger polypeptides that bind to sequences within gene regulatory
15 sequences. Moreover, this invention also relates to transgenic plants that comprise engineered zinc finger-containing peptides.

BACKGROUND OF THE INVENTION

The application of biotechnology to plants has yielded many agricultural gains. For example, biotechnology has been used to improve various properties of
20 plants such as resistance to pests, diseases and herbicides, and the improvement of various seed and fruit traits. Many further applications of plant biotechnology are anticipated and these include the modification of specific traits that may be of agronomic interest or of interest in the processing and use of plant-derived products. In many instances, this could be undertaken by the manipulation of endogenous
25 genes which encode these traits, however, a sophisticated means to achieve up- and down-regulation of such endogenous genes is, in many cases, not yet available. In addition, plants also hold great promise as biological "factories" for a variety of chemical products including enzymes and compounds for industrial and pharmaceutical use. However, it is expected that the continuous production of high
30 concentrations of gene products and compounds for such use may have deleterious

consequences for the host plant and consequently, more sophisticated mechanisms for expressing such genes are required.

Accordingly, gene switches are currently of interest for the control of timing and/or dosage of gene expression in plants. In particular, the development of gene switches that can be directed towards any gene in a plant chromosome is highly desirable.

SUMMARY OF THE INVENTION

The present invention seeks to overcome these difficulties by providing non-naturally occurring engineered zinc finger proteins to confer specificity on gene regulation for both target endogenous genes and transgenes. More specifically, the present invention can be used to regulate any gene in a plant.

Accordingly the present invention provides a method of regulating transcription in a plant cell by introducing an engineered zinc finger polypeptide into said plant cell which polypeptide binds to a target DNA and modulates transcription of a coding sequence which is operably linked to said target DNA.

Engineered zinc finger polypeptides have been used to regulate genes in mammalian cells. Choo et al. Nature 372:642-645 (1994); Pomerantz et al. Science 267:93-96 (1995); Liu et al. Proc. Natl. Acad. Sci. USA 94:5525-5530 (1997); Beerli et al. Proc. Natl. Acad. Sci. USA 95:14628-14633 (1998). However, only in the case of Choo et al. was the regulated gene integrated in a chromosome of the host mammalian cell. It is well recognized that the biology of mammalian and plant cells is very different and that each has evolved be very different at the structural, physiological, biochemical and molecular biological level. In the present invention, the inventors have shown for the first time that it is possible to regulate a gene in a plant using an engineered zinc finger protein. More specifically, the inventors have shown that, using an engineered zinc finger protein, a gene integrated in a plant chromosome can be regulated via binding of the engineered zinc finger protein to a target DNA sequence adjacent to a target gene.

The zinc finger proteins of the present invention can be used to up-regulate or down-regulate any gene, particularly any gene in a plant. By designing a zinc finger protein with a transactivating domain, the specific induction of an endogenous

gene can be accomplished bypassing any endogenous regulation of the targeted gene. Previously, the only available method was to introduce a transgene in another location of the genome under the regulation of a separate promoter. The zinc finger proteins of the present invention can also be used to down-regulate any gene in any plant. This has previously only been possible using techniques such as antisense, ribozymes and co-suppression, all of which are somewhat unpredictable. The zinc finger protein approach to down-regulation is highly potent and allows, for instance, the targeting of a specific member of a gene family without affecting the other members.

The term "engineered" means that the zinc finger peptide, polypeptide, protein or chimera protein has been generated or modified in vitro. Typically the zinc finger polypeptide is produced by deliberate mutagenesis, for example the substitution of one or more amino acid residues, either as part of a random mutagenesis procedure or by site-directed mutagenesis, or by selection from a library or libraries of mutated zinc finger peptides. Engineered zinc finger peptides for use in the invention can also be produced de novo using rational design strategies.

The term "introduced into" means that a procedure is performed on a plant, a plant part, or a plant cell such that the gene encoding a zinc finger protein is then present in the cell or cells. Examples of suitable procedures include the microinjection, bombardment, biolistic bombardment, *Agrobacterium* transformation, electroporation, transfection or other transformation or delivery techniques to cells of a nucleic acid construct capable of directing expression of the zinc finger polypeptide, or the zinc finger protein itself in the cell.

The term "target DNA sequence" means any nucleic acid sequence to which a zinc finger peptide is capable of binding. It is usually but not necessarily a DNA sequence within a plant chromosome, to which an engineered zinc finger peptide is capable of binding. A target DNA sequence will generally be associated with a target gene (see below) and the binding of the engineered zinc finger polypeptide to the DNA sequence will generally allow the up- or down-regulation of the associated gene. In one embodiment, the target DNA sequence is part of an endogenous genomic sequence. In another embodiment, the target DNA sequence and associated

coding sequence have been introduced into the cell or are heterologous to the cell. In many cases, a target DNA sequence will form part of a promoter or other transcription regulatory region such as an enhancer. In a most preferred embodiment, the target DNA sequence is a known sequence of a promoter from a plant gene of interest.

The term "target gene" means a gene or other coding sequence the expression of which can be affected using compositions and methods described in the present invention. A target gene may be an endogenous gene (i.e. one which is normally found in genome of the plant or plant cell) or a heterologous gene (i.e. one that does not normally exist in the genome of the plant or cell).

The term "heterologous to the cell" means not naturally existing in the cell but having been introduced into the cell. A heterologous sequence can include a modified sequence introduced at any chromosomal site, or which is not integrated into a chromosome, or which is introduced by homologous recombination such that it is present in the genome in the same position as the native allele.

In a highly preferred embodiment, the zinc finger polypeptide is fused to a biological effector domain. "Biological effector domain" means any polypeptide that has a biological function and includes enzymes and transcriptional regulatory domains or proteins, and additional sequence such as nuclear localization sequences. Preferably the zinc finger polypeptide is fused to a transcriptional activator domain or a transcriptional repressor domain.

In a further embodiment of the method of the invention, the plant cell is part of a plant or can be regenerated into a plant and the target DNA sequence is part of a regulatory sequence to which the target gene is operably linked.

The present invention also provides a plant cell comprising a polynucleotide encoding an engineered zinc finger protein and a target DNA sequence to which the zinc finger protein binds.

The present invention also provides a transgenic plant comprising a polynucleotide encoding an engineered zinc finger polypeptide and a target DNA sequence to which the zinc finger protein binds.

The present invention further provides a transgenic plant comprising a polynucleotide encoding an engineered zinc finger protein and a target DNA sequence that is within a plant chromosome.

5 The present invention further provides a transgenic plant comprising a polynucleotide encoding an engineered zinc finger polypeptide and a target DNA sequence within the sequence of a gene endogenous to the plant.

The present invention further provides a large number of zinc finger peptides capable of binding to every defined target sequence which is advantageously of 3 or 4 nucleotides. The peptides can be incorporated into zinc finger polypeptides, zinc
10 finger proteins and zinc finger chimera proteins.

The present invention further provides promoter systems for high level expression of a target gene. The promoter systems contain, 5'-3' operably linked, a target DNA sequence, a minimal promoter and at least one target gene.

The present invention further provides methods of inducing high level
15 expression of a target gene in a promoter system containing, 5'-3' operably linked, a target DNA sequence, a minimal promoter and at least one target gene. Methods of inducing high level expression include expressing the zinc finger-activator protein that specifically binds to the target DNA sequence.

BRIEF DESCRIPTION OF THE DRAWINGS

20 Figure 1 is a general depiction of plasmids described herein and their use in transforming plants with *Agrobacterium*.

Figure 2 depicts production of positive transgenic lines more specifically.

Figure 3 depicts plasmids pBA002 and pER8.

Figure 4 depicts the component parts and final constructs of reporter
25 constructs, and pZVE1.

Figure 5 depicts transient expression of TFIIIAZIFVP16/VP64 and activation of a luciferase reporter construct in onion peels by bombardment-mediated transformation. In Figure 5, a. is TFIIIAZIFVP16 + 4XBS-Luciferase, b. is TFIIIAZIFVP16 + 1XBS-Luciferase, c. is TFIIIAZIFVP64 + 4XBS-Luciferase, d. is
30 TFIIIAZIFVP16 + 4XBS-Luciferase, e. is TFIIIAZIFVP64 + 4XBS-Luciferase, f. is 4XBS-Luciferase, g. is 1XBS-Luciferase, and h. is KIN2-Luciferase.

Figure 6 depicts 17- β -Estradiol (estrogen) regulated expression of luciferase. In Figure 6, a. is pER8-TFIIIAZIFVP64 + pIXBSluciferase (with estrogen), b. is pIXBSluciferase (with estrogen), c. is pER8-TFIIIAZIFVP64 + pIXBSluciferase (without estrogen) and d. is pIXBSluciferase (without estrogen).

5 Figure 7 depicts 17- β -Estradiol regulated expression of GFP. In Figure 7, a. is pER8-TFIIIAZIFVP64 + p4XBSGFP and b. is pER8-TFIIIAZIFVP16 + p4XBSGFP

Figure 8 depicts luciferase expression by induction of pER8-TFIIIAZIFVP16 in T1 transgenic plant leaves containing the pBA4XBSLUC and pER8-
10 TFIIIAZIFVP16 + p4XBSLUC.

Figure 9 depicts expression of 1XBSGFP and 1XBMRFP in onion peels.

Figure 10 depicts the results of a Northern blot of T1 transgenic lines. In Figure 10, A is RNA from leaves containing pER8TFIIIAZIFVP16 and Pba4Xbsluc + Estradiol (24h) and B is RNA from leaves containing Pba4Xbsluc + Estradiol
15 (24h).

Figure 11 depicts transient luciferase expression of 1XBSLUC + S35TFIIIAZif VP16 in Arabidopsis seedlings by gold particle bombardment-mediated transformation.

DETAILED DESCRIPTION OF THE INVENTION

20 Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art (e.g., in cell culture, molecular genetics, nucleic acid chemistry, hybridization techniques and biochemistry). Standard techniques are used for molecular, genetic and biochemical methods (see generally, Sambrook et al. Molecular Cloning: A
25 Laboratory Manual, 2d ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. and Ausubel et al. Short Protocols in Molecular Biology (1999) 4th Ed, John Wiley & Sons, Inc.), chemical methods, pharmaceutical formulations and delivery and treatment of patients.

A. Zinc Finger Chimera Proteins

30 A zinc finger chimera protein is a transcription factor that comprises a DNA binding domain (comprising a number of zinc finger peptides) designed to bind

specifically to any DNA sequence and one or more further domains. Usually, a nuclear localization domain is attached to the DNA binding domain to direct the chimera to the nucleus. Generally, the zinc finger chimera protein also includes an effector domain that can be a transactivation or repression domain to regulate the expression of the target gene. Choo and Klug (1995) *Curr. Opin. Biotech.* 6:431-436; Choo and Klug (1997); Rebar and Pabo (1994) *Science* 263:671-673; and Jamieson et al. (1994) *Biochem.* 33:5689-5695. The zinc finger chimera protein may also preferably include other domains which may be advantageous within the context of the present invention. For example, DNA modifying domains (such as endonucleases and methylases) can be added to the DNA binding domain, conferring to the zinc finger chimera protein the ability to regulate expression of the target gene or specifically modify any DNA. Wu et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:344-348; Nahon and Raveh (1998); Smith et al. (1999); and Carroll et al. (1999). Zinc finger proteins of the Cys2-His2 class are preferred within the context of the present invention.

Zinc finger peptides are small protein domains that are able to recognize and bind a nucleic acid triplet, or an overlapping quadruplet, in a nucleic acid binding sequence. Preferably, there are 2 or more zinc finger peptides, for example 2, 3, 4, 5 or 6 zinc finger peptides, in each zinc finger polypeptide. Advantageously, there are 3 or more zinc finger polypeptides in each zinc finger polypeptide. All of the DNA binding residue positions of zinc finger peptides, as referred to herein, are numbered from the first residue in the α -helix of the peptide, ranging from +1 to +9. "-1" refers to the residue in the framework structure immediately preceding the α -helix in a Cys2-His2 zinc finger peptide. Residues referred to as "++" are residues present in an adjacent (C-terminal) peptide. Where there is no C-terminal adjacent peptide, "++" interactions do not operate.

Zinc finger polypeptides according to the present invention are engineered. That is, essentially "man-made." Typically, zinc finger polypeptides according to the invention are produced by mutagenesis techniques or designed using rational design techniques. Zinc finger polypeptides can also be selected from randomized libraries using screening procedures, such as those described below.

The present invention is in one aspect concerned with the production of what are essentially engineered DNA binding proteins. In these proteins, artificial analogues of amino acid residues may be used, to impart the proteins with desired properties or for other reasons. Thus, the term "amino acid residue", particularly in the context where "any amino acid residue" is referred to, means any sort of natural or artificial amino acid or amino acid analogue that may be employed in protein construction according to methods known in the art. Moreover, any specific amino acid residue referred to herein may be replaced by a functional analogue thereof, particularly an artificial functional analogue. The nomenclature used herein therefore specifically comprises within its scope functional analogues or mimetics of the defined amino acids.

The α -helix of a zinc finger peptide aligns antiparallel to the target nucleic acid strand, such that the primary nucleic acid sequence is arranged 3' to 5' in order to correspond with the N- terminal to C-terminal sequence of the zinc finger peptide. Since nucleic acid sequences are conventionally written 5' to 3', and amino acid sequences N-terminus to C-terminus, the result is that when a target nucleic acid sequence and a zinc finger peptide are aligned according to convention, the primary interaction of the zinc finger peptide is with the - strand of the nucleic acid sequence, since it is this strand which is aligned 3' to 5'. These conventions are followed in the nomenclature used herein. It should be noted, however, that in nature certain zinc finger peptides, such as zinc finger peptide 4 of the protein GLI, bind to the + strand of nucleic acid sequence. See Suzuki et al. (1994) Nucl. Acids Res. 22:3397-3405; and Pavletich and Pabo (1993) Science 261:1701-1707. The present invention encompasses incorporation of such zinc finger peptides into DNA binding molecules.

The present invention may preferably be integrated with the rules set forth for zinc finger polypeptide engineering in our copending European or PCT patent applications having publication numbers WO 98/53057, WO 98/53060, WO 98/53058, which describe improved techniques for designing zinc finger polypeptides capable of binding target DNA sequences. In combination with selection procedures, such as phage display, set forth for example in WO 96/06166,

these techniques enable the production of zinc finger polypeptides capable of recognising practically any desired sequence.

The structure of zinc finger peptide DNA binding motifs is to those in the art and defined in, for example, Miller et al. (1985) EMBO J. 4:1609-1614; Berg (1988) Proc. Natl. Acad. Sci. USA 85:99-102; and Lee et al. (1989) Science 245:635-637; see also International patent applications WO 96/06166 and WO 96/32475, corresponding to USSN 08/422,107, incorporated herein by reference.

In general, a preferred zinc finger peptide DNA binding motif has the structure:

10 (A) X₀₋₂ C X₁₋₅ C X₉₋₁₄ H X₃₋₆ H/C

where X is any amino acid residue, and the numbers in subscript indicate the possible numbers of residues represented by X.

In a preferred aspect of the present invention, zinc finger peptide DNA binding motifs may be represented as motifs having the following primary structure:

15 (B) X_a C X₂₋₄ C X₂₋₃ F X_c X X X X L X X H X X X_b H - linker
-1 1 2 3 4 5 6 7 8 9

wherein X (including X^a, X^b and X^c) is any amino acid residue. X_{2,4} and X_{2,3} refer to the presence of 2 or 4, or 2 or 3, amino acid residues, respectively. The Cys and His residues, which together co-ordinate the zinc metal atom, are marked in bold text and are usually invariant, as is the Leu residue at position +4 in the α -helix.

Modifications to this representation may occur or be effected without necessarily abolishing zinc finger peptide function, by insertion, mutation or deletion of amino acid residues. For example the second His residue may be replaced by Cys (Krizek et al. (1991) J. Am. Chem. Soc. 113:4518-4523) and that Leu at +4 can in some circumstances be replaced with Arg. The Phe residue before X_c may be replaced by any aromatic residue other than Trp. Moreover, experiments have shown that departure from the preferred structure and residue assignments for a zinc finger peptide are tolerated and may even prove beneficial in binding to certain nucleic acid sequences. Even taking this into account, however, the general structure involving an α -helix co-ordinated by a zinc atom which contacts four Cys or His residues, is not altered. As used herein, structures (A) and (B) above are

taken as an exemplary structure representing all zinc finger peptide structures of the Cys2-His2 type.

Preferably, X^a is $F/Y-X$ or $P-F/Y-X$. In this context, X is any amino acid residue. Preferably, in this context X is E, K, T or S. Less preferred but also
5 encompassed, are Q, V, A and P. The remaining amino acid residues remain possible.

Preferably, $X_{2,4}$ consists of two amino acid residues rather than four. The first of these may be any residue, but S, E, K, T, P and R are preferred. Advantageously, it is P or R. The second of these residues is preferably E, although
10 any residue may be used.

Preferably, X^b is T or I. Preferably, X^c is S or T.

Preferably, $X_{2,3}$ is G-K-A, G-K-C, G-K-S or G-K-G. However, departures from the preferred residues are possible, for example in the form of M-R-N or M-R.

Preferably, the linker is T-G-E-K or T-G-E-K-P.

15 In one approach to zinc finger engineering, the major nucleic acid binding interactions are assumed to occur with amino acid residues -1, +3 and +6. Amino acid residues +4 and +7 are largely invariant. The remaining residues may be essentially any amino acid residue. Preferably, position +9 is occupied by Arg or Lys. Advantageously, positions +1, +5 and +8 are not hydrophobic residues, that is
20 to say are not Phe, Trp or Tyr. Preferably, position ++2 is any residue, and preferably serine, save where its nature is dictated by its role as a ++2 residue for an N-terminal zinc finger peptide in the same nucleic acid binding molecule.

In a most preferred aspect, therefore, bringing together the above, the invention allows the definition of every residue in a zinc finger peptide DNA
25 binding motif which will bind specifically to a given target DNA triplet.

The code provided by the present invention is not entirely rigid; certain options are provided. For example, positions +1, +5 and +8 may be any amino acid residue, whilst other positions may have certain options: for example, the present rules provide that, for binding to a central T residue, any one of Ala, Ser or Val may
30 be used at +3. In its broadest sense, therefore, the present invention provides a very

large number of proteins that are capable of binding to every defined target DNA triplet.

Preferably, however, the number of possibilities may be significantly reduced. For example, the non-critical residues +1, +5 and +8 may be occupied by the residues Lys, Thr and Gln respectively as a default option. In the case of the other choices, for example, the first-given option may be employed as a default. Thus, the code according to the present invention allows the design of a single, defined polypeptide (a "default" polypeptide) which will bind to its target DNA triplet.

Accordingly, the zinc finger polypeptides of the present invention can be prepared using a method comprising the steps of: (a) selecting a model zinc finger peptide from the group consisting of naturally occurring zinc finger proteins and consensus zinc finger peptides; and (b) mutating at least one of positions -1, +3, +6 (and ++2) of the peptide.

In general, naturally occurring zinc finger proteins may be selected from those proteins for which the DNA binding specificity is known. For example, these may be the proteins for which a crystal structure has been resolved: namely Zif 268 (Elrod-Erickson et al. (1996) Structure 4:1171-1180), GLI (Pavletich and Pabo (1993) Science 261:1701-1707), Tramtrack (Fairall et al. (1993) Nature 366:483-487) and YY1 (Houbaviy et al. (1996) Proc. Natl. Acad. Sci. USA 93:13577-13582).

Although mutation of the DNA-contacting amino acid residues of the DNA binding domain of zinc finger peptides allows selection of peptides which bind to desired target nucleic acids, in a preferred embodiment residues which are outside the DNA-contacting region may be mutated. Mutations in such residues may affect the interaction between zinc finger peptides in a zinc finger polypeptide, and thus alter binding site specificity.

The naturally occurring zinc finger protein 2 in Zif 268 makes an excellent starting point from which to engineer a zinc finger polypeptide and is preferred.

Consensus zinc finger peptide structures may be prepared by comparing the sequences of known zinc finger peptides, irrespective of whether their binding

domain is known. Preferably, the consensus structure is selected from the group consisting of the consensus structure PYKCPECGKSFSQKSDLVKHQRTHTG (SEQ ID NO:1), and the consensus structure PYKCSECGKAFSQKSNLTRHQRIHTGEKP (SEQ ID NO:2).

- 5 The consensuses are derived from the consensus provided by Krizek et al. (1991) J. Am. Chem. Soc. 113:4518-4523 and from Jacobs (1993) Ph.D. thesis, University of Cambridge, UK. In both cases, the linker sequences described above for joining two zinc finger peptides together, namely TGEK or TGEKP can be formed on the ends of the consensus. Thus, a P may be removed where necessary,
10 or, in the case of the consensus terminating T G, E K (P) can be added.

The present invention provides methods of engineering and using zinc finger proteins in plants which zinc finger proteins are capable of binding to a target DNA sequence, and wherein binding to each base of the triplet by an α -helical zinc finger peptide DNA binding motif in the protein is determined as follows:

- 15 (a) if the 5' base in the triplet is G, then position +6 in the α -helix is Arg and/or position ++2 is Asp;
(b) if the 5' base in the triplet is A, then position +6 in the α -helix is Gln or Glu and ++2 is not Asp;
(c) if the 5' base in the triplet is T, then position +6 in the α -helix is Ser or Thr
20 and position ++2 is Asp; or position +6 is a hydrophobic amino acid other than Ala;
(d) if the 5' base in the triplet is C, then position +6 in the α -helix may be any amino acid, provided that position ++2 in the α -helix is not Asp;
(e) if the central base in the triplet is G, then position +3 in the α -helix is His;
(f) if the central base in the triplet is A, then position +3 in the α -helix is Asn;
25 (g) if the central base in the triplet is T, then position +3 in the α -helix is Ala, Ser, Ile, Leu, Thr or Val; provided that if it is Ala, then one of the residues at -1 or +6 is a small residue;
(h) if the central base in the triplet is 5-mC, then position +3 in the α -helix is Ala, Ser, Ile, Leu, Thr or Val; provided that if it is Ala, then one of the residues at -1
30 or +6 is a small residue;
(i) if the 3' base in the triplet is G, then position -1 in the α -helix is Arg;

- (j) if the 3' base in the triplet is A, then position -1 in the α -helix is Gln and position +2 is Ala;
- (k) if the 3' base in the triplet is T, then position -1 in the α -helix is Asn; or position -1 is Gln and position +2 is Ser; and
- 5 (l) if the 3' base in the triplet is C, then position -1 in the α -helix is Asp and Position +1 is Arg; where the central residue of a target triplet is C, the use of Asp at position +3 of a zinc finger polypeptide allows preferential binding to C over 5-mC.

The foregoing represents a set of rules which permits the design of a zinc finger peptide specific for any given target DNA sequence.

- 10 When the nucleic acid specificity of the model zinc finger peptide selected is known, the mutation of the peptide in order to modify its DNA binding specificity may be directed to residues known to affect binding to nucleotides at which the native and target DNAs differ. Otherwise, mutation of the model zinc finger peptides should be concentrated upon residues -1, +3, +6 and ++2 as provided for in
- 15 the foregoing rules.

In order to produce a zinc finger protein having improved binding, moreover, the rules provided by the present invention may be supplemented by physical or virtual modeling of the protein/DNA interface in order to assist in residue selection.

- 20 Methods for the production of libraries encoding randomised polypeptides are known in the art and may be applied in the present invention. Randomisation may be total, or partial; in the case of partial randomisation, the selected codons preferably encode options for amino acid residues as set forth in the rules above.

- 25 The invention encompasses library technology described in our copending International patent application WO 98/53057, incorporated herein by reference in its entirety. WO 98/53057 describes the production of zinc finger polypeptide libraries in which each individual zinc finger polypeptide comprises more than one, for example two or three, zinc finger peptides; and wherein within each polypeptide partial randomisation occurs in at least two zinc finger peptides.

- 30 This allows for the selection of the "overlap" specificity, wherein, within each triplet, the choice of residue for binding to the third nucleotide (read 3' to 5' on the + strand) is influenced by the residue present at position +2 on the subsequent

zinc finger peptide, which displays cross-strand specificity in binding. The selection of zinc finger polypeptides incorporating cross-strand specificity of adjacent zinc finger peptides enables the selection of nucleic acid binding proteins more quickly, and/or with a higher degree of specificity than is otherwise possible.

5 Zinc finger peptide binding motifs engineered for use in accordance with the present invention may be combined into nucleic acid binding proteins having a multiplicity of zinc finger peptides. Preferably, these chimeric proteins have at least two zinc finger peptides. In nature, zinc finger proteins commonly have at least three zinc finger peptides, although proteins containing two-zinc finger peptides
10 such as Tramtrack are known. The presence of at least three zinc finger peptides is preferred. Nucleic acid binding proteins may be constructed by joining the required zinc finger peptides end to end, N-terminus to C-terminus. Preferably, this is effected by joining together the relevant nucleic acid sequences encoding the zinc finger peptides to produce a composite nucleic acid coding sequence encoding the
15 entire zinc finger polypeptide.

A "leader" peptide may be added to the N-terminal zinc finger peptide. Preferably, the leader peptide is MAEEKP.

Zinc finger polypeptides comprising more than three zinc finger peptides, such as four, five, six, seven, eight or nine zinc finger peptides can also be used in
20 conjunction with the present invention. Linkers that are preferably used to link zinc finger peptides are described in co-pending patent applications GB 0013102.9, GB 0013103.7 and GB 0013104.5 filed on 30 May 2000. An example of a multiple zinc finger polypeptide described in this specification comprises zinc finger peptides 1-3 of TFIIIA and the three zinc finger peptides from Zif268 joined by zinc finger
25 peptide 4, including flanking sequences, of TFIIIA. We have called this zinc finger protein TFIIIAZif. Zinc finger peptide 4 of TFIIIA does not bind DNA but acts as a linker in between the two sets of zinc finger polypeptides that are involved in DNA recognition. Despite the fact that this zinc finger peptide does not make any base contacts within the major groove of the DNA, it is folded in the classical way, for
30 Cys2His2 zinc finger proteins, around a Zn(II) ion and is folded to contain an α helix within its structure (Nolte et al. 1998). However, other linkers can be used in

conjunction with the present invention to construct proteins comprising multiple zinc finger peptides.

B. Target Genes

5 Examples of target genes include any gene or coding region involved in any trait that may be of interest to a scientist, farmer or grower, a processor, or a consumer of a plant or plant product.

For example, genes involved in plant starch characteristics are useful target genes and the present invention can be used in conjunction with starch branching enzyme (for example) to generate corn plants which generate seed with super-
10 branching starch.

Genes involved in oil characteristics are useful target genes and the present invention can be used in conjunction with delta-12-desaturase (for example) to generate corn plants which generate seed with higher oleic and lower linoleic acid.

15 Genes involved in cotton fiber characteristics are useful target genes and the technology of the present invention can be used to modify the expression of such genes to improve traits such as fiber strength.

A further example is provided by the genes involved in the biosynthesis and catabolism of gibberellins. The gibberellins are a class of plant hormones involved in the determination of many plant traits including elongation growth,
20 bolting/flowering, leaf expansion, seed set, fruit size and dormancy. Accordingly, the regulation of genes involved in the biosynthesis and catabolism of gibberellins can be used to generate plants such as wheat, corn, sugar beet and sugar cane with improved traits. Phillips et al. (1995) Plant Physiol. 108:1049-1057; MacMillin et al. (1997) Plant Physiol. 113:1369-1377; Williams et al. (1998) Plant Physiol.
25 117:559-563; and Thomas et al. (1999) Proc. Natl. Acad. Sci. USA 96:4698-4703.

Genes involved in nitrogen metabolism (for example glutamine synthetase, asparagine synthetase, GOGAT, glutamate dehydrogenase) are useful target genes and the technology of the present invention can be used to modify the expression of such genes to improve nitrogen use efficiency in plants and thereby to reduce the
30 requirement for the application of inorganic fertilizers.

Genes involved in the biosynthesis of plant cell components such as cellulose and lignin can be targeted by the technology of the present invention to modify digestibility in crops such as corn which are used as silage.

Genes whose products are responsible for ripening (such as
5 polygalacturonase and ACC oxidase) are interesting target genes as the present invention can be used to modify ripening characteristics in plants such as tomato, avocado and banana.

Genes involved in the biosynthesis of volatile esters, important flavor compounds in fruits and vegetables, are equally interesting target genes and the
10 present invention can be used to improve such traits. Dudavera et al. (1996) Plant Cell 8:1137-1148; Dudavera et al. (1998) Plant J. 14:297-304; and Ross et al. (1999) Arch. Biochem. Biophys. 367:9-16.

Genes involved in the biosynthesis of plant-derived pharmaceutically important compounds are also potential target genes. Using the technology of the
15 present invention, the up-regulation of a rate-limiting step in the biosynthetic pathway of a pharmaceutically important compound results in production of higher levels of such compound in the plant.

Additionally, target genes used in conjunction with the present invention include genes encoding allergens such as the peanut allergens Arah1, Arah2 and
20 Arah3. Rabjohn et al. J. Clin. Invest. 103:535-542. Down-regulation of such genes using the technology of the present invention is expected to reduce the allergenicity of the transgenic peanuts.

Examples of heterologous target genes include genes which are introduced into a plant for the production of biodegradable plastic (for example) but which are
25 placed under the regulatory control of a zinc finger protein of the present invention.

C. Target DNA Sequences

Most commonly, target DNA sequences will be sequences associated with a target gene that is to be regulated by a zinc finger protein of the present invention. Target DNA sequences include not only sequences naturally associated with target
30 genes, but also other sequences which can be configured with a target gene to allow the up- or down regulation of such gene. For example, the known binding site of a

given zinc finger protein can be a target DNA sequence and, when operably linked to a target gene, will allow expression of the target gene to be regulated by the given zinc finger protein.

D. DNA Libraries

5 DNA sequences for use in screening methods to select DNA encoding zinc finger peptides and to provide target DNA sequences can be provided as a library of related sequences having homology to one another (as opposed to a genomic library, for example, obtained by cloning a large number of essentially unrelated sequences).

10 A library of DNA sequences can be used in at least two different ways. First, it can be used in a screen to identify zinc finger peptides that bind to a specific sequence. Second, it can be used to confirm the specificity of selected zinc finger peptides.

A DNA library is advantageously used to test the selectivity of a zinc finger peptide for nucleotide sequences of length N . Consequently, since there are four
15 different nucleotides that occur naturally in genomic DNA, the total number of sequences required to represent all possible base permutations for a sequence of length N is 4^N . N is an integer having a value of at least three. That is to say that the smallest library envisaged for testing binding of a zinc finger peptide to a nucleotide sequence where only one DNA triplet is varied, consists of 64 different sequences.
20 However, N can be any integer greater than or equal to 3 such as 4, 5, 6, 7, 8 or 9. Typically, N only needs to be three times the number of zinc finger peptides being tested, optionally including a few additional residues outside of the DNA binding site that can influence specificity. Thus, by way of example, to test the specificity of a protein comprising three zinc finger peptides, where all three peptides have been
25 engineered, it can be desirable to use a library where N is at least 9.

Libraries of DNA sequences can be screened using a number of different methods. For example, the DNA can be immobilised to beads and incubated with zinc finger peptides that are labeled with an affinity ligand such as biotin or
30 expressed on the surface of phage. Complexes between the DNA and zinc finger peptide can be selected by washing the beads to remove unbound peptides and then purifying the beads using the affinity ligand bound to the peptides to remove beads

that do not contain bound peptides. Any remaining beads should contain DNA/zinc finger peptide complexes. Individual beads can be selected and the identity of the DNA and zinc finger peptide determined. Other modifications to the technique include the use of detectable labels, for example fluorescently labeling the zinc finger peptides and sorting by zinc finger peptide bound-FACS beads.

In an alternative method, the DNA sequences in the library are immobilised at discrete positions on a solid substrate, such as a DNA chip, such that each different sequence is separated from other sequences on the solid substrate. Binding of zinc finger peptides is determined as described below and individual peptides isolated (which can be conveniently achieved by the use of phage display techniques). This technique can also be used as a second step after a zinc finger polypeptide has been selected by, for example, the bead method described above, to characterise fully the binding specificity of a selected zinc finger polypeptide, protein or chimera protein.

In a DNA library, it is generally not necessary or desirable for all positions to be randomised. Preferably only a subsequence of N bases of the complete DNA sequence is varied. The 4^N possible permutations of the DNA sequence of length N sequence are typically arranged in $4N$ sub-libraries, wherein for any one sub-library one base in the DNA sequence of length N is defined and the other $N-1$ bases are randomised. Thus in the case of a varied DNA triplet, there will be 12 sub-libraries.

As mentioned above, the nucleotide sequence of length N is generally part of a longer DNA molecule. However, the nucleotide sequence of length N typically occupies the same position within the longer molecule in each of the varied sequences even though the sequence of N itself can vary. The other sequences within the DNA molecule are generally the same throughout the library. Thus the library can be said to consist of a library of 4^N DNA molecules of the formula $R^1-(A/C/G/T)_N-R^2$, wherein R^1 and R^2 can be any nucleotide sequence.

Preferably, each sequence is also represented as a dilution/concentration series. Thus the immobilized DNA library can occupy $Z4^N$ discrete positions on the chip where Z is the number of different dilutions in the series and is an integer having a value of at least 2. The range of DNA concentrations for the dilution series is

typically in the order of 0.01 to 100 pmol cm⁻², preferably from 0.05 to 5 pmol cm⁻². The concentrations typically vary 10-fold, i.e. a series can consist of 0.01, 0.1, 1, 10 and 100 pmol cm⁻², but can vary, for example, by 2- or 5-fold.

The advantage of including the DNA sequences in a dilution series is that it is then possible to estimate K_ds for protein/DNA complexes using standard techniques such as the Kaleidagraph™ version 2.0 program (Abelback Software).

The DNA molecules in the library are at least partially double-stranded, in particular at least the nucleotide sequence of length N is double-stranded. Single stranded regions can be included, for example to assist in attaching the DNA library to a solid substrate.

Techniques for producing immobilized libraries of DNA molecules have been described in the art. Generally, most prior art methods described how to synthesize single-stranded nucleic acid molecule libraries, using for example masking techniques to build up various permutations of sequences at the various discrete positions on the solid substrate. U.S. Patent No. 5,837,832 (the '832 patent), describes an improved method for producing DNA arrays immobilized to silicon substrates based on very large scale integration technology. In particular, the '832 patent describes a strategy called "tiling" to synthesize specific sets of probes at spatially-defined locations on a substrate which can be used to produce the immobilized DNA libraries of the present invention. The '832 patent also provides references for earlier suitable techniques.

However, an important aspect of the present invention is that it relates to DNA binding proteins, zinc finger proteins that bind double-stranded DNA. Thus single-stranded nucleic acid molecule libraries using the prior art techniques referred to above will then need to be converted to double-stranded DNA libraries by synthesizing a complementary strand. An example of the conversion of single-stranded nucleic acid molecule libraries to double-stranded DNA libraries is given in Bulyk et al. (1999) Nature Biotechnol. 17:573-577. The technique described in Bulyk et al. (1999) typically requires the inclusion of a constant sequence in every member of the library (i.e. within R¹ or R² in the generic formula given above) to which a nucleotide primer is bound to act as a primer for second strand synthesis using a DNA polymerase and other appropriate reagents. If required, deoxynucleotide triphosphates (dNTPs) having

a detectable labeled can be included to allow the efficiency of second strand synthesis to be monitored. Also the detectable label can assist in detecting binding of zinc finger proteins when the immobilized DNA library is in use.

Alternatively, double-stranded DNA molecules can be synthesized off the solid substrate and each pre-formed sequence applied to a discrete position on the solid substrate. An example of such a method is to synthesize palindromic single-stranded nucleic acids. See U.S. Patent No. 5,556,752.

Thus DNA can typically be synthesized in situ on the surface of the substrate. However, DNA can also be printed directly onto the substrate using for example robotic devices equipped with either pins or piezo electric devices.

The library sequences are typically immobilized onto or in discrete regions of a solid substrate. The substrate can be porous to allow immobilization within the substrate or substantially non-porous, in which case the library sequences are typically immobilized on the surface of the substrate. The solid substrate can be made of any material to which polypeptides can bind, either directly or indirectly. Examples of suitable solid substrates include flat glass, silicon wafers, mica, ceramics and organic polymers such as plastics, including polystyrene and polymethacrylate. It can also be possible to use semi-permeable membranes such as nitrocellulose or nylon membranes, which are widely available. The semi-permeable membranes can be mounted on a more robust solid surface such as glass. The surfaces can optionally be coated with a layer of metal, such as gold, platinum or other transition metal. A particular example of a suitable solid substrate is the commercially available BiaCore™ chip (Pharmacia Biosensors).

Preferably, the solid substrate is generally a material having a rigid or semi-rigid surface. In preferred embodiments, at least one surface of the substrate will be substantially flat, although in some embodiments it can be desirable to physically separate synthesis regions for different polymers with, for example, raised regions or etched trenches. Preferably the solid substrate is not a microtiter plate or bead. It is also preferred that the solid substrate is suitable for the high density application of DNA sequences in discrete areas of typically from 50 to 100 μm , giving a density of 10000 to 40000 cm^{-2} .

The solid substrate is conveniently divided up into sections. This can be achieved by techniques such as photoetching, or by the application of hydrophobic inks, for example Teflon-based inks (Cel-line, USA). Discrete positions, in which each different member of the library is located can have any convenient shape, e.g.,

5 circular, rectangular, elliptical, wedge-shaped, etc.

Attachment of the library sequences to the substrate can be by covalent or non-covalent means. The library sequences can be attached to the substrate via a layer of molecules to which the library sequences bind. For example, the library sequences can be labeled with biotin and the substrate coated with avidin and/or

10 streptavidin. A convenient feature of using biotinylated library sequences is that the efficiency of coupling to the solid substrate can be determined easily. Since the library sequences can bind only poorly to some solid substrates, it is often necessary to provide a chemical interface between the solid substrate (such as in the case of glass) and the library sequences.

15 Examples of suitable chemical interfaces include hexaethylene glycol. Another example is the use of polylysine coated glass, the polylysine then being chemically modified using standard procedures to introduce an affinity ligand. Other methods for attaching molecules to the surfaces of solid substrate by the use of coupling agents are known in the art, see for example WO98/49557.

20 Binding of zinc finger polypeptides to the immobilized DNA library can be determined by a variety of means such as changes in the optical characteristics of the bound DNA (i.e. by the use of ethidium bromide) or by the use of labeled zinc finger polypeptides, such as epitope tagged zinc finger polypeptides or zinc finger polypeptides labeled with fluorophores such as green fluorescent protein (GFP). Other

25 detection techniques that do not require the use of labels include optical techniques such as optoacoustics, reflectometry, ellipsometry and surface plasma resonance (SPR). See, WO97/49989.

Binding of epitope tagged zinc finger polypeptides is typically assessed by immunological detection techniques where the primary or secondary antibody

30 comprises a detectable label. A preferred detectable label is one that emits light, such as a fluorophore, for example phycoerythrin.

The complete DNA library is typically read at the same time by charged coupled device (CCD) camera or confocal imaging system. Alternatively, the DNA library can be placed for detection in a suitable apparatus that can move in an X-Y direction, such as a plate reader. In this way, the change in characteristics for each discrete position can be measured automatically by computer controlled movement of the array to place each discrete element in turn in line with the detection means.

E. Nucleic Acid Vectors Encoding Zinc Finger Proteins

Polynucleotides encoding zinc finger proteins for use in the invention can be incorporated into a recombinant replicable vector. The vector can be used to replicate the nucleic acid in a compatible host cell and the vector can be recovered from the host cell. Suitable host cells for this purpose include bacteria such as *Escherichia coli*, yeast and eukaryotic cell lines. Alternatively, the vector can be used to deliver DNA capable of expressing the zinc finger protein to host plant cells.

Preferably, a vector containing a polynucleotide encoding a zinc finger protein according to the invention is operably linked to a control sequence that is capable of providing for the expression of the coding sequence by the host cell, i.e. the vector is an expression vector. The term "operably linked" means that the components described are in a relationship permitting them to function in their intended manner. A regulatory sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

The control sequences can be modified, for example by the addition of further transcriptional regulatory elements to make the level of transcription directed by the control sequences more responsive to transcriptional modulators.

Vectors of the invention can be transformed or transfected into a suitable host cell as described below to provide for expression of a protein of the invention.

The vectors can be for example, plasmid or virus vectors provided with an origin of replication, optionally a promoter for the expression of the polynucleotide and optionally a regulator of the promoter. The vectors can contain one or more selectable marker genes and these will vary depending on the system used, but are known to those of skill in the art. Vectors can be used, for example, to transfect or

transform a host cell. US Patent Nos. 6,130,066; 6,162,966; 6,147,280; 6,147,279; 6,147,278; 6,147,277; and 5,792,926 involve vectors and cite numerous documents concerning making and using vectors.

Control sequences operably linked to sequences encoding the protein of the invention include promoters/enhancers and other expression regulation signals such as terminators. These control sequences can be selected to be compatible with the host cell in which the expression vector is designed to be used. The term promoter is well-known in the art and encompasses nucleic acid regions ranging in size and complexity from minimal promoters to promoters including upstream elements and enhancers.

For the purpose of expressing a zinc finger protein in plant cells, the promoter is typically selected from promoters that function in plant cells, prokaryotic promoters and promoters functional in other eukaryotic cells can be used in addition to promoters obtained from plants. Typically, the promoter is derived from viral or plant gene sequences. For example, the promoter can be derived from the genome of a cell in which expression is to occur. With respect to plant promoters, they can be promoters that function in a ubiquitous manner or, alternatively, a tissue-specific manner. Tissue-specific promoters specific for different tissues of the plant are particularly preferred. Examples are provided below. Tissue-specific expression can be used to confine expression of the protein to a cell type or tissue/organ of interest. Promoters can also be used that respond to specific stimuli, for example promoters that are responsive to plant hormones. Viral promoters can also be used, for example the CaMV 35S promoter well known in the art.

It can also be advantageous for the promoters to be inducible so that the levels of expression of the heterologous gene can be regulated during the life-time of the cell. Inducible means that the levels of expression obtained using the promoter can be regulated. Inducible expression allows the researcher to control when expression of the polypeptides takes place.

In addition, any of the promoters can be modified by the addition of further regulatory sequences, for example enhancer sequences. Chimeric promoters can also contain sequence elements from two or more different promoters.

Advantageously, a plant expression vector encoding a zinc finger protein according to the invention can comprise a locus control region (LCR). LCRs are capable of directing high-level integration site independent expression of transgenes integrated into host cell chromatin. Stief et al. (1989) Nature 341:343; and Lang et al. (1991) Nucl. Acids Res. 19:5851-5856..

According to the invention, the zinc finger protein constructs of the invention are expressed in plant cells under the control of transcriptional regulatory sequences that are known to function in plants. The regulatory sequences selected will depend on the required temporal and spatial expression pattern of the zinc finger protein in the host plant. Many plant promoters have been characterized and would be suitable for use in conjunction with the invention. By way of illustration, some examples are provided below:

A large number of promoters direct expression in specific tissues and organs (e.g. roots, leaves, flowers) or in cell types (e.g. leaf epidermal cells, leaf mesophyll cells, root cortex cells). For example, the maize PEPC promoter from the phosphoenol carboxylase gene (Hudspeth and Grula (1989) Plant Mol. Biol. 12:579-589) is green tissue-specific; the *trpA* gene promoter is pith cell-specific (WO 93/07278 to Ciba-Geigy); the TA29 promoter is pollen-specific. Mariani et al. (1990) Nature 347:737-741; and Mariani et al. (1992) Nature 357:384-387.

Other promoters direct transcription under conditions of presence of light or absence of light or in a circadian manner. For example, the GS2 promoter described by Edwards and Coruzzi (1989) Plant Cell 1:241-248 is induced by light, whereas the AS1 promoter described by Tsai and Coruzzi (1990) EMBO J. 9:323-332 is expressed only in conditions of darkness.

Other promoters are wound-inducible and typically direct transcription not just upon wound induction, but also at the sites of pathogen infection. Examples are described by Xu et al. (1993) Plant Mol. Biol. 22:573-588; Logemann et al. (1989) Plant Cell 1:151-158; and Firek et al. (1993) Plant Mol. Biol. 22:129-142.

Further plant promoters of interest are the bronze promoter (Ralston et al. (1988) Genetics 119:185-197 and Genbank Accession No. X07937.1) which directs expression of UDPglucose flavanoid glycosyl-transferase in maize, the patatin-1

gene promoter (Jefferson et al. (1990) Plant Mol. Biol. 14:995-1006) that contains sequences capable of directing tuber-specific expression, and the phenylalanine ammonia lyase promoter (Bevan et al. (1989) EMBO J. 8:1899-1906) thought to be involved in responses to mechanical wounding and normal development of the xylem and phloem.

A number of constitutive promoters can be used in plants. These include the Cauliflower Mosaic Virus (CaMV) 35S promoter (US 5,352,605 and US 5,322,938, both to Monsanto) including minimal promoters (such as the CaMV 35S-90, CaMV 35S-46 and bronze minimal promoter) linked to other regulatory sequences, the rice actin promoter (McElroy et al. (1991) Mol. Gen. Genet. 231:150-160), and the maize and sunflower ubiquitin promoters. Christensen et al. (1989) Plant Mol. Biol. 12:619-632; and Binet et al. (1991) Plant Science 79:87-94).

A further promoter of interest is the inducible promoter described by Aoyama and Chua (1997) Plant J. 11:605-612; and Zou and Chua (2000) Curr. Op. Biotech. 11:146-151. By using this inducible promoter system, transgenic lines can be established which carry a zinc finger chimera protein but express it only after addition of an inducer. Thus the genes encoding the zinc finger proteins of the present invention can be expressed in response to the inducer allowing the dose or level of zinc finger protein in the cell or plant to be adjusted to a desired amount.

Using promoters that direct transcription in the plant species of interest, the gene encoding the zinc finger protein of the invention can be expressed in the required cell or tissue types. For example, if it is the intention to utilize the zinc finger protein to regulate a gene in a specific cell or tissue type, then the appropriate promoter can be used to direct expression of the zinc finger protein gene construct.

An appropriate terminator of transcription is fused downstream of the gene encoding the selected zinc finger protein. Any of a number of available terminators can be used in conjunction with the invention. Examples of transcriptional terminator sequences known to function in plants include, without limitation, the nopaline synthase terminator found in the pBI vectors (Clontech catalog 1993/1994), the E9 terminator from the *rbcS* gene, and the *tm1* terminator from CaMV.

A number of sequences found within the transcriptional unit are known to enhance gene expression and these can be used within the context of the current invention. Such sequences include intron sequences which, particularly in monocotyledon cells, are known to enhance expression. Both intron 1 of the maize *Adh1* gene and the intron from the maize *bronze1* gene have been found to be effective in enhancing gene expression in maize cells (Callis et al. (1987) Genes Develop. 1:1183-1200) and intron sequences are frequently incorporated into plant transformation vectors, typically within the non-translated leader.

A number of virus-derived non-translated leader sequences have been found to enhance expression, especially in dicotyledon cells. Examples include the "Ω" leader sequence of Tobacco Mosaic Virus, and similar leader sequences of Maize Chlorotic Mottle Virus and Alfalfa Mosaic Virus. Gallie et al. (1987) Nucl. Acids Res. 15:8693-8711; and Shuzeski et al. (1990) Plant Mol. Biol. 15:65-79.

The zinc finger proteins of the current invention are targeted to the cell nucleus so that they are able to interact with host cell DNA and bind to the appropriate DNA target in the nucleus and regulate transcription. To effect this, a Nuclear Localization Sequence (NLS) is incorporated in frame with the expressible zinc finger protein gene construct. The NLS can be fused either 5' or 3' to the sequence encoding the zinc finger protein.

The NLS of the wild-type Simian Virus 40 Large T-Antigen (Kalderon et al. (1984) Cell 37:801-813; and Markland et al. (1987) Mol. Cell. Biol. 7:4255-4265) is an appropriate NLS and provides an effective nuclear localization mechanism in plants. van der Krol et al. (1991) Plant Cell 3:667-675. However, several alternative NLSs are known in the art and can be used instead of the SV40 NLS sequence. These include the NLSs of TGA-1A and TGA-1B. van der Krol et al. (1991).

A variety of transformation vectors are available for plant transformation and the zinc finger protein-encoding genes of the invention can be used in conjunction with any such vectors. The selection of vector will depend on the preferred transformation technique and the plant species that is to be transformed. For certain target species, different selectable markers can be preferred.

For *Agrobacterium*-mediated transformation, binary vectors or vectors carrying at least one T-DNA border sequence are suitable. A number of vectors are available including pBIN19 (Bevan (1984) Nucl. Acids Res. 12:8711-8721), the pBI series of vectors, and pCIB10 and derivatives thereof. Rothstein et al. (1987) Gene
5 53:153-161; and WO 95/33818.

Binary vector constructs prepared for *Agrobacterium* transformation are introduced into an appropriate strain of *Agrobacterium tumefaciens* (for example, LBA 4044 or GV 3101) either by triparental mating or direct transformation. Bevan (1984); and Höfgen and Willmitzer (1988) Nucl. Acids Res. 16:9877.

10 For transformation which is not *Agrobacterium*-mediated (i.e. direct gene transfer), any vector is suitable and linear DNA containing only the construct of interest can be preferred. Direct gene transfer can be undertaken using a single DNA species or multiple DNA species (cotransformation; Schroder et al. (1986) Biotechnol. 4:1093-1096).

15 Construction of vectors according to the invention employs conventional ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to generate the vectors required. If desired, analysis to confirm correct sequences in the constructed plasmids is performed in a known fashion. Suitable methods for constructing expression vectors, preparing in vitro
20 transcripts, introducing DNA into host cells, and performing analyses for assessing DNA binding protein expression and function are known to those skilled in the art. Gene presence, amplification and/or expression can be measured directly in a sample, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA, dot blotting (DNA or RNA analysis), or in
25 situ hybridization, using an appropriately labeled probe which can be based on a sequence provided herein. Those skilled in the art will readily envisage how these methods can be modified, if desired.

F. Construction of Transgenic Plants Expressing Zinc Finger Proteins

Heterologous DNA can be introduced into plant host cells by any method
30 known in the art, such as electroporation, particle bombardment or *A. tumefaciens* mediated transfer. Although specific protocols can vary from species to species,

transformation techniques are well known in the art for most commercial plant species. Dicotyledonous species include, for example, *Arabidopsis*, tobacco, cotton, tomato, canola and oilseed rape, poplar, potato, sunflower, soybean, pea, cauliflower, various bean species and lentil. Monocotyledonous species include, for example, maize, wheat, barley, oat and rice.

In the case of dicotyledonous species, *Agrobacterium*-mediated transformation is generally a preferred technique as it has broad application to many dicotyledon species and is generally very efficient. *Agrobacterium*-mediated transformation generally involves the co-cultivation of *Agrobacterium* with explants from the plant and follows procedures and protocols known in the art. Transformed tissue is generally regenerated on medium carrying the appropriate selectable marker. Protocols are known in the art for many dicotyledonous crops including (for example) cotton, tomato, canola and oilseed rape, poplar, potato, sunflower, tobacco and soybean. See for example EP 0 317 511, EP 0 249 432, WO 87/07299, US 5,795,855.

In addition to *Agrobacterium*-mediated transformation, various other techniques can be applied to dicotyledons. These include polyethylene glycol (PEG) and electroporation-mediated transformation of protoplasts, and microinjection. Potrykus et al. (1985) Mol. Gen. Genet. 199:169-177; Reich et al. (1986) Biotechnol. 4:1001-1004; and Klein et al. (1987) Nature 327:70-73. As with *Agrobacterium*-mediated transformation, transformed tissue is generally regenerated on medium carrying the appropriate selectable marker using standard techniques known in the art.

Although *Agrobacterium*-mediated transformation has been applied successfully to monocotyledonous species such as rice and maize and protocols for these approaches are available in the art, the most widely used transformation techniques for monocotyledons remain particle bombardment, and PEG and electroporation-mediated transformation of protoplasts.

In the case of maize, techniques are available for transformation using particle bombardment. Gordon-Kamm et al. (1990) Plant Cell 2:603-618; Fromm et al. (1990) Biotechnol. 8:833-839; and Koziel et al. (1993) Biotechnol. 11:194-200.

The preferred method is the use of biolistics using, for instance, gold or tungsten. Suitable methods are known in the art and described, for instance, in US Patent Nos. 5,489,520 and 5,550,318. See also, Potrykus (1990) *Bio/Technol.* 8:535-542; and Finnegan et al. (1994) *Bio/Technol.* 12:883-888.

5 In the case of rice, protoplast-mediated transformation for both *Japonica*- and *Indica*-types has been described (Zhang et al. (1988) *Plant Cell Rep.* 7:379-384; Shimamoto et al. *Nature* 338:274-277; and Datta et al. (1990) *Biotechnol.* 8:736-740) and both types are also routinely transformable using particle bombardment. Christou et al. (1991) *Biotechnol.* 9:957-962.

10 In the case of wheat, transformation by particle bombardment has been described for both type C long-term regenerable callus (Vasil et al. (1992) *Biotechnol.* 10:667-674) and immature embryos and immature embryo-derived callus (Vasil et al. (1993) *Biotechnol.* 11:1553-1558; and Weeks et al. (1993) *Plant Physiol.* 102:1077-1084). A further technique is described in published patent
15 applications WO 94/13822 and WO 95/33818.

Transformation of plant cells is normally undertaken with a selectable marker that can provide resistance to an antibiotic or to a herbicide. Selectable markers that are routinely used in transformation include the *nptII* gene which confers resistance to kanamycin (Messing & Viera (1982) *Gene* 19:259-268; and Bevan et al. (1983)
20 *Nature* 304:184-187), the *bar* gene which confers resistance to the herbicide phosphinothricin (White et al. (1990) *Nucl. Acids Res.* 18:1062; and Spencer et al. (1990) *Theor. Appl. Genet.* 79:625-631), the *hph* gene which confers resistance to the antibiotic hygromycin (Blochliger and Diggelmann (1984) *Mol. Cell. Biol.* 4:2929-2931), and the *dhfr* gene which confers resistance to methotrexate (Bourouis
25 et al. (1983) *EMBO J.* 2:1099-1104). More recently, a number of selection systems have been developed which do not rely of selection for resistance to antibiotic or herbicide. These include the inducible isopentyl transferase system described by Kunkel et al. (1999) *Nature Biotechnol.* 17:916-919. A further such system is the red pigment produced under the control of the R-locus in maize. This pigment can
30 be detected by culturing cells on a solid support containing nutrient media capable of supporting growth at this stage, incubating the cells at, e.g., 18°C and greater than

180 μ Em-2s-1, and selecting cells from pigmented colonies (visible aggregates of cells). These cells can be cultured further, either in suspension or on solid media.

G. Optimizing Gene Expression in Plants

The zinc finger protein gene constructs of the invention are suitable for expression in a variety of different organisms. However, to enhance the efficiency of expression it may be necessary to modify the nucleotide sequence encoding the zinc finger protein to account for different frequencies of codon usage in different host organisms. Hence it is preferable that the sequences to be introduced into organisms, such as plants, conform to preferred codon usage in the host organism.

In general, high gene expression in plants is best achieved from codon sequences that have a GC content of at least 35% and preferably more than 45%. This is thought to be because the existence of ATTTA motifs destabilize mRNAs and the existence of AATAAA motifs can cause inappropriate polyadenylation, resulting in truncation of transcription. Murray et al. (1989) (Nucl. Acids Res. 17:477-498) have shown that even within plants, monocotyledonous and dicotyledonous species have differing preferences for codon usage, with monocotyledonous species generally preferring GC richer sequences. Thus, in order to achieve optimal high level expression in plants, gene sequences can be altered to accommodate such preferences in codon usage in such a manner that the codons encoded by the DNA are not changed.

Plants also have a preference for certain nucleotides adjacent to the ATG encoding the initiating methionine and for most efficient translation, these nucleotides can be modified. To facilitate translation in plant cells, it is preferable to insert, immediately upstream of the ATG coding the initiating methionine of the gene to be expressed, a "plant translational initiation context sequence." A variety of sequences can be inserted at this position. These include the sequence 5'-AAGGAGATATAACAATG-3' (SEQ ID NO:3) (Prasher et al. (1992) Gene 111:229-233; and Chalfie et al. (1992) Science 263:802-805), the sequence 5'-GTCGACCATG-3' (SEQ ID NO:4) (Clontech 1993/1994 catalog, page 210), and the sequence 5'-TAAACAATG-3'. Joshi et al. (1987) Nucl. Acids Res. 15:6643-6653. For any particular plant species, a survey of natural sequences available in any

databank (e.g. GenBank) can be undertaken to determine preferred "plant translational initiation context sequences" on a species-by-species basis.

Any changes that are made to the coding sequence can be made using techniques that are well known in the art and include site directed mutagenesis, PCR, and synthetic gene construction. Such methods are described in published patent applications EP 0 385 962, EP 0 359 472 and WO 93/07278. Well-known protocols for transient expression in plants can be used to check the expression of modified genes before their transfer to plants by transformation. Typically, transient expression is undertaken with an expression vector having a selectable marker gene, and DNA is delivered using techniques such as electroporation or particle bombardment into various target cell or tissue types.

H. Zinc Finger Protein Regulation of Gene Expression in vivo in Plants

The present invention provides a method of regulating gene expression in a plant using an engineered zinc finger protein. The present invention also provides a method for regulating gene expression in a plant using a promoter system in conjunction with a zinc finger protein.

Thus, zinc finger proteins such as those designed or selected as described above are useful in switching or modulating gene expression in plants, in particular with respect to agricultural biotechnology applications as described herein.

Usually, the zinc finger polypeptides will be fused to a transcriptional effector domain to activate or repress transcription from a gene downstream of a zinc finger DNA binding sequence. Additionally, a fusion polypeptide comprising a zinc finger polypeptide and a DNA cleavage domain can be used to regulate gene expression by specific recognition and cleavage of a nucleic acid sequence. In addition, zinc finger proteins capable of differentiating between U and T can be used to preferentially target RNA or DNA, as required. Furthermore, zinc finger proteins can be engineered to recognize methylated nucleotide sequences as described in patent publication WO 99/47656 and these zinc finger proteins are especially useful to bind to methylated target DNA sequences. When such zinc finger proteins are used in conjunction with a transcriptional activator, they can be used to induce expression of such a gene that may otherwise be silenced by methylation.

Thus zinc finger proteins according to the invention will typically require the presence of a transcriptional effector domain, such as an activation domain or a repressor domain. Examples of transcriptional activation domains include the VP16 and VP64 transactivation domains of Herpes Simplex Virus. Alternative
5 transactivation domains are various and include the maize C1 transactivation domain sequence (Sainz et al. (1997) Mol. Cell. Biol. 17:115-22) and P1 (Goff et al. (1992) Genes Dev. 6:864-75; and Estruch et al. (1994) Nucl. Acids Res. 22:3983-89) and a number of other domains that have been reported from plants. Estruch et al. (1994).

Instead of incorporating a transactivator of gene expression, a repressor of
10 gene expression can be fused to the gene encoding the zinc finger protein and used to down-regulate the expression of a gene contiguous with or incorporating the zinc finger DNA target sequence. Such repressors are known in the art and include, for example, the KRAB-A domain (Moosmann et al. (1997) Biol. Chem. 378:669-677) the *engrailed* domain (Han et al. (1993) EMBO J. 12:2723-2733) and the *snag*
15 domain (Grimes et al. (1996) Mol. Cell. Biol. 16:6263-6272). These can be used alone or in combination to down-regulate gene expression.

Another possible application discussed above is the use of zinc finger polypeptides fused to nucleic acid cleavage moieties, such as the catalytic domain of a restriction enzyme, to produce a restriction enzyme capable of cleaving only target
20 DNA of a specific sequence. Kim et al. (1996) Proc. Natl. Acad. Sci. USA 93:1156-1160. Using such approaches, different zinc finger polypeptide domains can be used to create restriction enzymes with any desired recognition nucleotide sequence. Preferably, expression of these zinc finger polypeptide-enzyme fusion proteins is inducible. Enzymes other than those that cleave nucleic acids can also be
25 used for a variety of purposes.

The target gene can be endogenous or heterologous to the genome of the cell, for example fused to a heterologous coding sequence. However, in either case it will be operably linked to a target DNA sequence, such as a target DNA sequence described above, to which a zinc finger protein according to the invention binds.
30 The zinc finger polypeptide is typically expressed from a DNA construct present in

the host cell which also contains the target DNA sequence. The DNA construct is preferably stably integrated into the genome of the host cell, but this is not essential.

Thus a host plant cell according to the invention comprises a target gene, a target DNA sequence and a construct capable of directing expression of the zinc
5 finger protein in the cell.

Suitable constructs for expressing the zinc finger protein are known in the art and are described in section E above. The coding sequence can be expressed constitutively or be regulated. Expression can be ubiquitous or tissue-specific. Suitable regulatory sequences are known in the art and are also described in section
10 E above. Thus the DNA construct will comprise a nucleic acid sequence encoding a zinc finger protein operably linked to a regulatory sequence capable of directing expression of the zinc finger protein in a host cell.

It can also be desirable to use target DNA sequences that include sequences that bind or overlap with transcriptional regulatory proteins, such as transactivators.
15 Preferably the transcriptional regulatory proteins are endogenous to the cell.

Techniques for introducing nucleic acid constructs into plant cells are known in the art and many are described both in section F.

"Transgenic" in the present context denotes organisms and more especially plants in which one or more cells receive a recombinant DNA molecule. Typically
20 the transgene introduced will be transferred to the next generation which is also thus denoted "transgenic."

The transgene introduced into the organism is preferably derived from a species foreign to the recipient animal (i.e., "heterologous"), but the transgene can also be foreign only to the particular individual recipient, or genetic information
25 already possessed by the recipient. In the last case, the transgene can be differently expressed than the native gene.

"Operably linked" refers to polynucleotide sequences that are necessary to effect the expression of coding and non-coding sequences to which they are properly ligated. The nature of such control sequences differs depending upon the host
30 organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence; in eukaryotes,

generally, such control sequences include promoters and a transcription termination sequence. The term "control sequences" is intended to include, at a minimum, components whose presence can influence expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

Thus, a polynucleotide construct for use in the present invention, to introduce a nucleotide sequence encoding a zinc finger protein into the genome of a multicellular organism, typically comprises a nucleotide sequence encoding the zinc finger protein operably linked to a regulatory sequence capable of directing expression of the coding sequence. In addition the polynucleotide construct can comprise flanking sequences homologous to the host cell organism genome to aid in integration.

The invention further provides promoter systems for high level expression of a target gene. The promoter systems contain, 5'-3' operably linked, a target DNA sequence, a minimal promoter and at least one target gene.

The present invention further provides methods of inducing high level expression of a target gene in a promoter system containing, 5'-3' operably linked, a target DNA sequence, a minimal promoter and at least one target gene. Methods of inducing high level expression include expressing the zinc finger-activator protein that specifically binds to the target DNA sequence.

The promoter systems contain any target DNA sequence known in the art such as those provided herein, any minimal promoter known in the art such as those provided herein and any target gene known in the art such as those described herein. A zinc finger protein capable of binding to the target DNA sequence is fused to a transcriptional activator and operably linked to a promoter that functions in a plant and used to transform a plant as described herein. Expression of the zinc finger chimera protein in a cell where the promoter system is present leads to high level expression of the target gene because the zinc finger chimera protein activates target gene expression via binding to the target DNA sequence.

In one embodiment, the promoter system comprises the TFIIIAZifVP16 zinc finger construct which is used to up-regulate the target gene having the 27 bp

binding site (or multimers thereof) of the TFIIIAZif zinc finger protein linked to the -46 CaMV minimal promoter operably linked to the target gene. As shown in the Examples presented herein, this promoter system results in high level expression of a target DNA sequence

- 5 The present invention will now be described by way of the following examples, which are illustrative only and non-limiting. The Examples show that a zinc finger chimera protein can be expressed in plants and recognizes a target DNA sequence in a plant genome. Secondly, the Examples show that zinc finger chimera proteins containing a transactivating domain can activate the expression of a target
- 10 gene in plants in a manner similar to zinc finger protein regulated gene expression in animal cells. Using this principle and the consensus methods described herein, zinc finger proteins can be designed to interact with specific target DNA sequences to either activate or repress the expression of target genes.

EXAMPLES

15 Materials and Methods

Gene Construction and Cloning.

- In general, procedures and materials are in accordance with guidance given in Sambrook et al. Molecular Cloning. A Laboratory Manual, Cold Spring Harbor, 1989. The gene encoding the Zif268 zinc finger peptides (residues 333-420) is
- 20 assembled from 8 overlapping synthetic oligonucleotides (see Choo and Klug (1994)), giving *SfiI* and *NotI* overhangs. The genes encoding zinc finger peptides of the phage library are synthesized from 4 oligonucleotides by directional end to end ligation using 3 short complementary linkers, and amplified by PCR from the single strand using forward and backward primers which contain sites for *NotI* and *SfiI*
- 25 respectively. Backward PCR primers in addition introduce Met-Ala-Glu as the first three amino acid residues of the zinc finger peptides, and these are followed by the residues of the wild type or library zinc finger peptides as required. Cloning overhangs were produced by digestion with *SfiI* and *NotI* where necessary. Fragments were ligated to 1 µg similarly prepared Fd-Tet-SN vector. This is a
- 30 derivative of fd-tet-DOG1 (Hoogenboom et al. (1991) Nucl. Acids Res. 19:4133-4137) in which a section of the pelB leader and a restriction site for the

enzyme *Sfi*I (underlined) have been added by site-directed mutagenesis using the oligonucleotide:

5'CTCCTGCAGTTGGACCTGTGCCATGGCCGGCTGGGCGCATAGAATGG
AACAACTAAAGC3' (SEQ ID NO:5)

5 which anneals in the region of the polylinker. Electrocompetent DH5α cells were transformed with recombinant vector in 200 ng aliquots, grown for 1 hour in 2xTY medium with 1% glucose, and plated on TYE containing 15 µg/ml tetracycline and 1% glucose.

The zinc finger chimera protein used for this first set of experiments is a fusion protein that comprises 4 domains. First, 4 zinc finger peptides of TFIIIA were linked through a spacer region to 3 zinc finger peptides of Zif268 and this construct is denoted TFIIIAZif. Choo and Klug (1997) Curr. Opin. Str. Biol. 7:117-125; Pavletich and Pabo (1991) Science 252:809-817; Elrod-Erickson et al. (1996) Structure 4:1171-1180; and Elrod-Erickson et al. (1998) Structure 6:451-464. This
15 designed zinc finger polypeptide is able to recognize specifically a DNA sequence of 27 base pair (bp). Second, a nucleic acid sequence encoding an NLS region rich in basic amino acid residues that directs the chimera to the nucleus was operably linked to the construct. Third, a transactivation domain from the Herpes Simplex Virus (HSV) VP16 or VP64 that is a tetramer of the minimal VP16 domain was operably
20 linked to the construct. This region activates gene expression. Last, the 9E10 region that corresponds to the myc domain for the specific antibody recognition of the expressed protein in plants was operably linked to the construct (Figure 3).

The reporter construct consists of a DNA monomer of the minimal binding site of 27 bp or tetramer of the minimal region that is recognized specifically by the
25 zinc finger polypeptide domain. This sequence is attached 5' to the 46 bp of the CaMV 35S minimal promoter. Downstream of the promoter the coding sequence of the luciferase or green fluorescent protein (GFP) genes were operably linked as reporter genes. The luciferase from *Photinus pyralis* catalyzes the ATP/oxygen-dependent oxidization of substrate luciferin which produces the emission of light
30 (bioluminescence) and the GFP fluoresces under blue light. At the 3' end of the construct the pea *rbcS-E9* polyadenylation sequence was operably linked (Figure 3).

The zinc finger peptide phage display library of the present invention contains amino acid randomisations in putative base-contacting positions from the second and third zinc finger peptides of the three-finger DNA binding domain of Zif268, and contains members that bind DNA of the sequence XXXXXGGCG where X is any base. Further details of the library used can be found in WO 98/53057.

Example 1

Generation of Transgenic Plants Expressing a Zinc Finger Protein Fused to a Transactivation Domain

To investigate the utility of heterologous zinc finger proteins for the regulation of plant genes, a synthetic zinc finger protein was designed and introduced into transgenic *A. thaliana* under the control of a promoter capable of expression in a plant as described below. A second construct comprising the zinc finger DNA binding sequence fused upstream of the GFP reporter gene was also introduced into transgenic *A. thaliana* as described in Example 2. Crossing the two transgenic lines produced progeny plants carrying both constructs in which the GFP reporter gene was expressed demonstrating transactivation of the gene by the zinc finger protein.

Using conventional cloning techniques, the following constructs were made as *XbaI-BamHI* fragments in the cloning vector pcDNA3.1 (Invitrogen).

pTFIIIAZifVP16

pTFIIIAZifVP16 comprises a fusion of four zinc finger peptide domains of the zinc finger protein TFIIIA fused to the three zinc finger peptides of the zinc finger protein Zif268. The TFIIIA-derived sequence was fused in frame to the translational initiation sequence ATG. The 7 amino acid residue NLS of the wild-type SV40 Large T-Antigen was fused to the 3' end of the Zif268 sequence, and the VP16 transactivation sequence was fused downstream of the NLS. In addition, a 30 bp sequence from the *c-myc* gene was introduced downstream of the VP16 domain as a "tag" to facilitate cellular localization studies of the transgene. While this was experimentally useful, the presence of this tag is not required for the activation (or repression) of gene expression via zinc finger proteins.

pTFIIIAZifVP16 is shown in Figure 4 as an XbaI-BamHI fragment. The translation initiating ATG is located at position 15 and is double underlined. Zinc fingers peptides 1 to 4 of TFIIIA extend from position 18 to position 416. Zinc finger peptide 4 (positions 308–416) does not bind DNA within the target DNA sequence, but instead serves to separate the first three zinc finger peptides of TFIIIA from Zif268 which is located at positions 417–689. The NLS is located at positions 701–722, the VP16 transactivation domain from positions 723–956, and the *c-myc* tag from positions 957–986. This is followed by the translational terminator TAA.

pTFIIIAZifVP64

pTFIIIAZifVP64 is similar to pTFIIIAZifVP16 except that the VP64 transactivation sequence replaces the VP16 sequence of pTFIIIAZifVP16.

pTFIIIAZifVP64 is shown in Figure 4 as an XbaI-BamHI fragment. Locations within this sequence are as for pTFIIIAZifVP16 except that the VP64 domain is located at position 723–908 and the *c-myc* tag from positions 909–938.

The target DNA sequence of the TFIIIAZif protein contains the DNA recognition sites for zinc finger peptides 1–3 of TFIIIA and the three zinc finger peptides of Zif 268. These are the DNA sequences GGATGGGAGAC (SEQ ID NO:6) and GCGTGGGCGT (SEQ ID NO:7), respectively. The six base pair sequence GTACCT is a spacer region of DNA that separates the two binding sites and the nucleotide composition of the DNA spacer appears to have no effect on

binding of the protein. Therefore, this or other structured linkers could be used with other DNA spacers of different length and sequence.

The amino acid sequence of zinc finger peptide 4 of TFIIIA, including the flanking sequences as used in the composite protein of the invention, is

5 NIKICVYVCHFENCGKAFFKKHNQLKVHQFSHTQQLP (SEQ ID NO:8).

The nucleotide sequence of zinc finger domain 4 of TFIIIA, including the flanking sequences, is

AACATCAAGATCTGCGTCTATGTGTGCCATTTTGAGAACTGTGGCAAAGC
ATTCAAGAAACACAATCAATTAAAGGTTTCATCAGTTCAGTCACACACAG

10 CAGCTGCCG (SEQ ID NO:9).

Using conventional cloning techniques, the sequence 5'-

AAGGAGATATAACA-3' (SEQ ID NO:10) is introduced upstream of the

translational initiating ATG of both pTFIIIAZifVP16 and pTFIIIAZifVP64. This sequence incorporates a plant translational initiation context sequence to facilitate

15 translation in plant cells. Prasher et al. (1992) Gene 111:229-233; and Chalfie et al. (1992) Science 263:802-805.

The final constructs were transferred to the plant binary vector pBIN121 between the CaMV 35S promoter and the nopaline synthase terminator sequence.

This transfer was effected using the *Xba*I site of pBIN121. The binary constructs

20 thus derived were then introduced into *A. tumefaciens* (strain LBA 4044 or GV 3101) either by triparental mating or direct transformation.

Next, *A. thaliana* are transformed with *Agrobacterium* containing the binary vector construct using conventional transformation techniques. For example, using vacuum infiltration (e.g. Bechtold et al. CR Acad. Sci. Paris 316:1194-1199; and

25 Bent et al. (1994) Science 265:1856-1860), transformation can be undertaken essentially as follows. Seeds of *Arabidopsis* are planted on top of cheesecloth

covered soil and allowed to grow at a final density of 1 per square inch under conditions of 16 hours light/8 hours dark. After 4-6 weeks, plants are ready to

infiltrate. An overnight liquid culture of *Agrobacterium* carrying the appropriate

30 construct is grown up at 28°C and used to inoculate a fresh 500ml culture. This

culture is grown to an OD₆₀₀ of at least 2.0, after which the cells are harvested by

centrifugation and resuspended in 1 liter of infiltration medium (1 liter prepared to contain: 2.2 g MS Salts, 1 X B5 vitamins, 50g sucrose, 0.5 g MES pH 5.7, 0.044 μ M benzylaminopurine, 200 L Silwet μ L-77 (OSI Specialty)). To vacuum infiltrate, pots are inverted into the infiltration medium and placed into a vacuum oven at room temperature. Infiltration is allowed to proceed for 5 mins at 400mm Hg. After releasing the vacuum, the pot is removed and laid it on its side and covered with Saran™ wrap. The cover is removed the next day and the plant stood upright. Seeds harvested from infiltrated plants are surface sterilized and selected on appropriate medium. Vernalization is undertaken for two nights at around 4°C. Plates are then transferred to a plant growth chamber. After about 7 days, transformants are visible and are transferred to soil and grown to maturity.

Transgenic plants are grown to maturity. They appear phenotypically normal and are selfed to homozygosity using standard techniques involving crossing and germination of progeny on appropriate concentration of antibiotic.

Transgenic plant lines carrying the TFIIIAZifVP16 construct are designated At-TFIIIAZifVP16 and transgenic plant lines carrying the TFIIIAZifVP64 construct are designated At-TFIIIAZifVP64.

Example 2

Generation of Transgenic Plants Carrying a

Green Fluorescent Protein Reporter Gene

A reporter plasmid is constructed which incorporates the target DNA sequence of the TFIIIAZifVP16 and TFIIIAZifVP64 zinc finger proteins described above upstream of the GFP reporter gene.

This sequence is incorporated in single copy immediately upstream of the CaMV 35S -90 or -46 minimal promoter to which the GFP gene is fused.

The resultant plasmid, designated pTFIIIAZif-UAS/GFP, is transferred to the plant binary vector pBIN121 replacing the CaMV 35S promoter. This construct is then transferred to *A. tumefaciens* and subsequently transferred to *A. thaliana* as described above. Transgenic plants carrying the construct are designated At-TFIIIAZif-UAS/GFP.

Example 3

Use of Zinc Finger Proteins to

Up-Regulate a Transgene in a Plant

To assess whether the zinc finger constructs TFIIIAZifVP16 and
5 TFIIIAZifVP64 are able to transactivate gene expression *in planta*, *Arabidopsis* lines
At-TFIIIAZifVP16 and *At*-TFIIIAZifVP64 are crossed to *At*-TFIIIAZif-UAS/GFP.
The progeny of such a cross yield plants that carry the reporter construct TFIIIAZif-
UAS/GFP together with either the zinc finger protein construct TFIIIAZifVP16 or
the zinc finger protein construct TFIIIAZifVP64.

10 Plants are screened for GFP expression using an inverted fluorescence
microscope (Leitz DM-IL) fitted with a filter set (Leitz-D excitation BP 355-425,
dichronic 455, emission LP 460) suitable for the main 395 nm excitation and 509 nm
emission peaks of GFP.

In each case, the zinc finger protein construct is able to transactivate gene
15 expression demonstrating the utility of heterologous zinc finger proteins for the
regulation of plant genes.

Example 4

Generation of Transgenic Plants Expressing

a Zinc Finger Polypeptide Fused to a Plant Transactivation Domain

20 The constructs pTFIIIAZifVP16 and pTFIIIAZifVP64 utilize the VP16 and
VP64 transactivation domains of Herpes Simplex Virus to activate gene expression.
Alternative transactivation domains are various and include the C1 transactivation
domain sequence (from maize; see Goff et al. (1991) Genes Dev. 5:298-309; and
Goff et al. (1992) Genes Dev. 6:864-875), and a number of other domains that have
25 been reported from plants. Estruch et al. (1994) Nucl. Acids Res. 22:3983-3989.

Construct pTFIIAZifC1 is made as described above for pTFIIIAZifVP16 and
pTFIIIAZifVP64 except the VP16/VP64 activation domains are replaced with the
C1 transactivation domain sequence

A transgenic *Arabidopsis* line, designated *At*-TFIIAZifC1, is produced as
30 described above in Example 2 and crossed with *At*-TFIIIAZif-UAS/GFP. The

progeny of such crosses yield plants that carry the reporter construct TFIIIAZif-UAS/GFP together with either the zinc finger protein construct TFIIIAZifC1.

Plants are screened for GFP expression using an inverted fluorescence microscope (Leitz DM-IL) fitted with a filter set (Leitz-D excitation BP 355-425, 5 dichronic 455, emission LP 460) suitable for the main 395 nm excitation and 509 nm emission peaks of GFP.

Example 5

Regulation of an Endogenous Plant

Gene – UDP glucose flavanoid glucosyl-transferase (UGFT).

10 To determine whether a suitably configured zinc finger protein could be used to regulate gene transcription from an endogenous gene in a plant, the maize UDP glucose flavanoid glucosyl-transferase (UGFT) gene (the Bronze1 gene) was selected as the target gene. UFGT is involved in anthocyanin biosynthesis. A number of wild type alleles have been identified including Bz-W22 that conditions a 15 purple phenotype in the maize seed and plant. The Bronze locus has been the subject of extensive genetic research because its phenotype is easy to score and its expression is tissue-specific and varied (for example aleurone, anthers, husks, cob and roots). The complete sequence of Bz-W22 including upstream regulatory sequences has been determined. Ralston et al. Genetics 119:185-197. A number of 20 sequence motifs that bind transcriptional regulatory proteins have been identified within the Bronze promoter including sequences homologous to consensus binding sites for the myb- and myc-like proteins. Roth et al. Plant Cell 3:317-325. Identification of a zinc finger protein that binds to the Bronze promoter

25 The first step is to carry out a screen for zinc finger proteins that bind to a selected region of the Bronze promoter. A region is chosen just upstream of the AT rich block located at between -88 and -80, which has been shown to be critical for Bz1 expression. Roth et al. supra. Using the methodology for zinc finger protein engineering described in detail elsewhere in this specification, three and six zinc finger peptide units were selected to bind the region upstream of -88.

Using conventional cloning techniques and in a similar manner to Example 1, the constructs pZifBz23C1, pZifBz23(6)VP16, and pZifBz23(6)KRAB are made in cloning vector pcDNA3.1 (Invitrogen).

pZifBz23C1 comprises the three zinc finger peptides of the zinc finger protein clone ZifBz23 fused in frame to the translational initiation sequence ATG. The 7 amino acid residue NLS of the wild-type SV40 Large T-Antigen is fused to the 3' end of the ZifBz23 sequence, and the C1 transactivation sequence is fused downstream of the NLS. In addition, 30 bp sequence from the *c-myc* gene is introduced downstream of the VP16 domain as a "tag" to facilitate cellular localization studies of the transgene. pZifBz23(6)VP16 is similar to pZifBz23C1, but instead comprises the six zinc finger peptide sequence and the VP16 activation domain instead of C1; and pZifBz23(6)KRAB is similar to pZifBz23(6)VP16 except that it comprises the KRAB repression domain instead of VP16.

The coding sequences of pZifBz23C1, pZifBz23(6)VP16 and pZifBz23(6)KRAB are transferred to a plant expression vector suitable for use in maize protoplasts and maize plants, the coding sequence being under the control of a constitutive CaMV 35S promoter. The resulting plasmids are termed pTMBz23C1, pZifBz23(6)VP16 and pZifBz23(6)KRAB. The vector also contains a hygromycin resistance gene or a bar gene for selection purposes.

A suspension culture of maize cells is prepared from calli derived from embryos obtained from inbred W22 maize stocks and other maize stocks grown to flowering in a greenhouse and self pollinated using essentially the protocol described in EP-A-332104 (Examples 40 and 41). The suspension culture is then used to prepare protoplasts using essentially the protocol described in EP-A-332104 (Example 42).

Protoplasts are resuspended in 0.2 M mannitol, 0.1% w/v MES, 72 mM NaCl, 70 mM CaCl₂, 2.5 mM KCl, 2.5 mM glucose with the pH is adjusted to 5.8 with KOH, at a density of about 2×10^6 per ml. 1 ml of the protoplast suspension is then aliquotted into plastic electroporation cuvettes and 10 µg of linearized pTMBz23 added. Electroporation is carried out as described in EP-A-332104

(Example 57). Protoplasts are cultured following transformation at a density of 2×10^6 per ml in KM-8p medium with no solidifying agent added.

Measurements of the levels UFGT expression are made using colorimetry and/or biochemical detection methods such as Northern blots or the enzyme activity assays described by Dooner and Nelson (1977) Proc. Natl. Acad. Sci. USA 74:5623-5627. Comparison is made with mock treated protoplasts transformed with a vector only control.

In addition to, analyzing expression of UFGT in transformed protoplasts, intact maize plants are recovered from transformed protoplasts and the extent of UFGT expression determined. Suitable protocols for growing maize plants from transformed protoplasts are known in the art. Electroporated protoplasts are resuspended in Km-8p medium containing 1.2% w/v Seaplaque agarose and 1 mg/l 2,4-D. Once the gel has set, protoplasts in agarose are placed in the dark at 26°C. After 14 days, colonies arise from the protoplasts. The agarose containing the colonies is transferred to the surface of a 9 cm diameter Petri dish containing 30 ml of N6 medium (EP-A-332,104) containing 2,4-D solidified with 0.24% Gelrite®. 100 mg/l hygromycin B is also added to select for transformed cells. The callus is cultured further in the dark at 26°C and callus pieces subcultured every two weeks onto fresh solid medium. Pieces of callus can be analyzed for the presence of the pTMBz23 construct and/or UFGT expression determined. In addition, transgenic maize plants are also generated by particle bombardment as described in Gordon-Kamm et al. (1990) Plant Cell 2:603-618; Fromm et al. (1990) Biotechnol. 8:833-839; and Koziel et al. (1993) Biotechnol. 11:194-200.

Corn plants are regenerated as described in Example 47 of EP-A-332,104. Plantlets appear in 4 to 8 weeks. When 2 cm tall, plantlets are transferred to ON6 medium (EP-A-332,104) in GA7 containers and roots form in 2 to 4 weeks. After transfer to peat pots, plants soon become established and can then be treated as normal corn plants.

Plantlets and plants are assayed for UFGT expression as described above.

Example 6

Cloning of Zinc Finger Chimera Protein in the Expression Vectors

In the following Examples, zinc finger chimera proteins specially designed for binding to specific target DNA sequences in plants, were engineered with an effector domain (transactivator or repressor) and expressed in plants using a series of inducible gene expression system XVE or ZVE1. Aoyama and Chua (1997); and Zou and Chua (2000). Plant transformation was performed using standard procedures utilizing *Agrobacterium*.

The general strategy is outlined in Figure 1 and more specifically in Figure 2.

The zinc finger chimera proteins were expressed either under the control of an estrogen receptor-based chemical-inducible system (binary vector pER8, Zuo et al. (2000) Plant J. 24:265-273) or the constitutive CaMV 35S promoter (binary vector pBa002, Hajdukiewicz et al. (1994) Plant Mol. Biol. 25:989-994). Plasmid pBa002 was digested with MluI and SpeI (New England BioLabs, MA). The coding region of the genes encoding zinc finger chimera proteins (VP16 and VP64) were engineered by PCR to have a MluI restriction site at the 5' end and a SpeI site at the 3' end. The sequences of the forward and the reverse primers were 5'CCACGCGTCCATGGGAGAGAAGGCGCTGCCGGTGG 3' (SEQ ID NO: 11) and 5'CCACTAGTCCTTACAGATCTTCTTCAGAAATAAGTTTTGTGCC 3' (SEQ ID NO:12), respectively. The PCR-amplified DNA fragment was digested with MluI and SpeI, gel purified using the Qiaquick Gel extraction protocol (Qiagen, Valencia, CA), and ligated into plasmid pBa002 using T4 DNA ligase (New England Biolabs, MA). A clone for each construct was verified by restriction analysis.

All constructs were introduced into *A. tumefaciens* strain ABI. Aoyama and Chua (1997). Similar procedures were used to clone the zinc finger chimera proteins (VP16 and VP64) into the AscI and SpeI sites. The coding region of each gene encoding a zinc finger chimera protein (VP16 and VP64) was engineered by PCR to have an AscI restriction site at the 5' end and a SpeI site at the 3' end.

Reporter construct

- pKL+1 plasmid was used for the construction of reporter plasmids. Foster and Chua (1999) Plant J. 17:363-372. pKL+1 plasmid contains a minimal promoter region from CaMV 35S promoter (−46 nucleotides) upstream of the luciferase coding sequence, that is terminated by pea *rbcS*-E9 polyadenylation sequence. The pKL+1 plasmid was digested with XbaI and HindIII (New England BioLabs, MA). A tetramer of the target DNA sequence of the zinc finger chimera protein was engineered by annealing two complementary oligos. The oligos were designed to have an XbaI restriction site at the 5' end and a HindIII site at the 3' end. The sequence of the sense and anti-sense strand primers were
- 5' CCTCTAGATCGGTCTCCCATCCAGGTACACGCCCACGCAAGTCGGTCTC
CCATCCAGGTACACGCCCACGCAAGTCGGTCTCCCATCCAGGTACACGC
CCACGCAAGTCGGTCTCCCATCCAGGTACACGCCCACGCAAGAAGCTTC
C 3' (SEQ ID NO:13) and
- 5' GGAAGCTTCTTGCGTGGGCGTGTACCTGGATGGGAGACCGACTTGCGT
GGGCGTGTACCTGGATGGGAGACCGACTTGCGTGGGCGTGTACCTGGAT
GGGAGACCGACTTGCGTGGGCGTGTACCTGGATGGGAGACCGATCTAGA
GG 3' (SEQ ID NO:14), respectively. The oligos were heated to 100°C for 5 min in TE (10 mM Tris-HCl pH8.5, 1mM EDTA) solution containing 500 mM NaCl and cooled to room temperature. The annealed oligos were isolated from an agarose gel using the Qiaquick Gel extraction protocol (Qiagen, Valencia, CA). The double stranded DNA fragment was digested with XbaI and HindIII, gel purified using the Qiaquick Gel extraction protocol (Qiagen, Valencia, CA), and ligated into the pKL+1 plasmid using T4 DNA ligase (New England Biolabs, MA). A clone for each construct was verified by restriction analysis. Similar procedures were used to engineer a single binding site reporter construct except that the oligos used contained an XbaI restriction site at the 5' end and a HindIII site at the 3' end. The sequences of the forward and the reverse primers were
- 5' CCAGATCTGGTCTCCCATCCAGGTACACGCCCACGCAAGATCTCC 3'
(SEQ ID NO:15) and

5'GGAGATCTTGCGTGGGCGTGTACCTGGATGGGAGACCAGATCTCGG3'
(SEQ ID NO:16), respectively.

For the versions of pKL+1 plasmid containing the GFP and RFP (red fluorescent protein), pKL+1 plasmid was digested with NcoI and KpnI for GFP and
5 SalI and KpnI for RFP (New England BioLabs, MA). The coding region of GFP was engineered by PCR to have an NcoI restriction site at the 5' end and an EcoRI site at the 3' end. The sequences of the forward and the reverse primers were
5'CCCCATGGTGAGCAAGGGCGAGGAGCTGTTCAACC 3' (SEQ ID NO:17)
and 5'CCGAATTCTTACTTGTACAGCTCGTCCATGCCGAG 3' (SEQ ID
10 NO:18), respectively. The coding region of RFP was engineered by PCR to have a SalI restriction site at the 5' end and an EcoRI site at the 3' end. The sequences of the forward and the reverse primers were
5'CCCTCGAGCGGGGTACCGCGGGCCCGGG3' (SEQ ID NO:19) and
5'CAGTTGGAATTCTAGAGTCGCGGCCGCTAC3' (SEQ ID NO:20),
15 respectively.

Example 7

Construction of the pZVE Plasmids

The new binary transformation plasmid pER12 (Zuo et al. 2000) was modified by replacing the LexA DNA binding domain with the zinc finger target
20 DNA sequence (TFIIIAZif, Figure 3). The coding region of the VP16-estrogen receptor was engineered by PCR to have XhoI restriction sites at both the 5' and 3' end. The sequences of the forward and the reverse primers were
5'CCGCTCGAGGCCCCCGACCGATGTCAGCCTGGGGGA3' (SEQ ID NO:21) and 5'CCG CTCGAGTATTAATTTGAGAATGAACAAAAAGGACC3'
25 (SEQ ID NO:22), respectively. The PCR-amplified DNA fragment was digested with XhoI, gel purified using the Qiaquick Gel extraction protocol (Qiagen, Valencia, CA), and ligated into the pTFIIIAZif plasmid (previously digested with XhoI) using T4 DNA ligase (New England Biolabs, MA). A clone was verified by restriction analysis and sequencing (pTFIIIAZif- VP16-ER). The coding region of
30 the TFIIIAZif-VP16-estrogen receptor fusion gene was engineered by PCR to have an AseI restriction site at both the 5' and 3' ends. The sequences of the forward and

the reverse primers were

5'GCCATTAATCGGAATGGGAGAGAAGGCGCTGCCGGTGG3' (SEQ ID NO:23) and 5'GCCTATTAATTTGAGAATGAACAAAAAGGACC3' (SEQ ID NO:24), respectively. pER12 plasmid was digested with AseI to removed the LexA-VP16-ER region. The PCR-amplified DNA fragment was digested with AseI, gel purified using the Qiaquick Gel extraction protocol (Qiagen, Valencia, CA), and ligated into the AseI-digested pER12 plasmid using T4 DNA ligase (New England Biolabs, MA). A clone was verified by restriction analysis and sequencing. The resulting plasmid (PER8 TFIIIAZif, Figure 3) containing the zinc finger target DNA sequence was digested with SalI to remove the hexamer of the LexA binding site.

The DNA fragment containing the plasmid was gel purified using the Qiaquick Gel extraction protocol (Qiagen, Valencia, CA), and ligated to a double strand tetramer of zinc finger target DNA sequences (previously digested with SalI) using T4 DNA ligase (New England Biolabs, MA). A clone was verified by restriction analysis and sequencing. The coding region of the GFP was engineered by PCR to have an XhoI restriction site at both the 5' and 3' ends and cloned in the multiple cloning site of the vector (Figure 4).

Example 8

Plant Transformation

Arabidopsis thaliana ecotype *Landsberg erecta* were transformed with *Agrobacterium* using the vacuum infiltration procedure according to Bent et al. (1994) Science 265:1856-1860. Seeds collected from the vacuum infiltrated plants were surface-sterilized by treatment with a solution of 1.5% sodium hypochlorite/0.01% Tween-20 (Sigma, St Louis MO, USA) for 10 min and washed three times with sterile water. The sterilized seeds were then resuspended in 0.1% agarose and sown in Petri dishes containing A medium (full-strength Murashige and Skoog salts, pH 5.7, 1 % sucrose, solidified with 0.8% Bactoagar, Gibco BRL, Grand Island, NY) and 20 µg/ml hygromycin B (Sigma, St. Louis, MO). The plated seeds were vernalized for 4 d and then transferred to a growth chamber maintained at 22°C under long day conditions (16h light/8h dark). Transgenic T1 seedlings were selected on a plate containing hygromycin (20 mg/ml) and after 2-3 weeks of growth

the presence of the transgene was confirmed by PCR analysis. The results are presented in Table 1.

Table 1

<u>Constructs</u>	<u>Onion/Trans</u>	<u>Trans/Plants</u>
35sVP16/4xBSLUC	Positive	---
35sVP16/1xBSLUC	Positive	---
35sVP64/4xBSLUC	Positive	---
35sVP64/1xBSLUC	Positive	---
ERVP16/4xBSLUC	Positive(+)	T1 gene
ERVP64/4xBSLUC	Positive(+)	T1 gene
4Xbsluc	---	T1 gene

5

Example 9

Estrogen Treatments

β -17-Estradiol (Sigma, St Louis MO) was dissolved in dimethylsulfoxide (DMSO) to make a 100 mM stock solution. The solution was stored at -20°C . To monitor transgene expression, transgenic seeds were surface sterilized and sown in Petri dishes as described above. After vernalization at 4°C for 4 days, the plates were incubated for two weeks in a growth chamber maintained at 22°C under long day (16hr light/8hr dark) conditions. Seedlings were removed from the plates and grown for 2 days in a hydroponic system containing liquid A medium. Fresh medium containing either β -17-Estradiol (30 μM) or no β -17-Estradiol was added, plants were removed at the designated time points, and then washed and frozen in liquid nitrogen. For the vector control transgenic lines similar conditions were used and the experiments were performed in parallel.

RNA analysis

Total RNA was isolated from seedlings and adult plants using the Qiagen RNA purification kit (Qiagen, Valencia, CA). RNA gel blot analysis was carried out according to the method described by Ausubel (1994). Each lane contained 10 μg of total RNA. The gene encoding the zinc finger polypeptide, luciferase and 18S rDNA fragments were obtained by PCR amplification with Pfu polymerase as described

above. Fragments were purified using the Qiaquick Gel extraction protocol (Qiagen, Valencia, CA). All DNA fragments were labeled with ^{32}P -dCTP and ^{32}P -dATP by random priming (Amersham, Arlington Heights, IL). Hybridization signals were quantified using the Phosphoimager STORM system (Molecular Dynamics) and the data analyzed with the Image Quant v1.1 program.

Light microscope and luciferase imaging

The GFP and RFP fluorescent microscopy analysis was done using an Axioskop (Zeiss, Germany) according to methods described by Mayer et al. (1993). Onion peels and 3 week old seedlings were sprayed with 2.5 mM luciferin (Promega) containing 0.005% Triton X-100 and the luciferase activity monitored by photon counting. Video images (5min) were captured in gravity mode using a intensifying CCD camera and coupling MethaMorph software (Universal Imaging Corporation PA). The results are presented in Figures 5-9.

Example 10

Results of Examples 6-9

In this series of experiments the zinc finger chimera protein was cloned into the plant expression vector pBA002, placing its expression under the regulation of the constitutively expressed CaMV S35 promoter (see Figure 3). The zinc finger chimera protein construct and the reporter construct were co-transformed into onion peels and *Arabidopsis* seedlings by biolistic bombardment with DNA-coated gold particles (BioRad, Oxford.UK). Transient expression of luciferase was recorded after 24 hr expression using an Imaging camera system. The results are shown in Figures 5 and 11. The results reveal that expression of either zinc finger chimera protein containing VP16 or VP64 was able to induce strong expression of luciferase.

When tested with a reporter vector containing one binding site, the zinc finger chimera protein containing VP16 induced luciferase activity. In the absence of the zinc finger chimera protein the reporter vectors with either a single copy or tetramer of the minimal binding site alone produced only background levels of luciferase expression (Figures 5 and 11).

In a second series of experiments, we used the XVE inducible system to express the zinc finger chimera protein and the reporter construct in onion peels and

young *Arabidopsis* seedlings. The XVE system is an estrogen receptor- based chemical-inducible system for expression of genes in transgenic plants (Figure 3).

Briefly, a chimeric transcriptional activator XVE is a fusion protein of the DNA binding domain from the bacterial repressor LexA (X), the acidic
5 transactivator domain VP16 (V) and the regulatory region of the human estrogen receptor (E, ER). The gene encoding the zinc finger chimera protein was cloned into the multiple cloning site (MCS) of the XVE vector downstream of eight copies of the LexA operator. The XVE was expressed by constitutive promoter G10-90 and in the presence of β -17 Estradiol the activated XVE binds the LexA operator, inducing
10 the expression of the zinc finger chimera protein. Similar experiments were done using the zinc finger chimera protein XVE construct, and the reporter constructs were transformed into two onion peels, one of which was sprayed with β -17-Estradiol. The transient expression of luciferase was recorded after 24 hr, and only the onion peels sprayed with β -17 Estradiol displayed strong expression of
15 luciferase (Figure 6).

Transformation of onion peels with the reporter vector alone containing a tetramer of the minimal binding site alone did not produce any luciferase expression with or without of β -17 Estradiol. Similar results were obtained with a reporter construct containing GFP (Figure 7). In addition, *Arabidopsis* plants were
20 transformed with pBA4Xluciferase and pBA4Xluciferase + pER8TFIIIAZifVP16. Only transgenic (T1) lines leaves containing the zinc finger chimera protein XVE construct (pER8TFIIIAZifVP16) were able to express luciferase with the addition of β -17 Estradiol (Figures 8 and 10). The results were confirmed by Northern blot analysis which showed expression of the zinc finger chimera protein and luciferase
25 in the total RNA extracted from leaves.

In a final experiment of the second series we tested the specificity of zinc finger chimera protein by comparing two reporter constructs containing a 27bp binding sequence upstream of GFP and the other with a single-mutated one 27bp sequence binding sequence upstream of the RFP cDNA. The single mutation in the
30 27bp target DNA sequence reduced the binding affinity of a zinc finger chimera protein by 10-20 fold. Preliminary experiments done in onion peels showed a 20

fold difference in the percentage of transformed cells (Figure 9). These constructs can be co-transformed with the inducible promoter system and a zinc finger chimera protein into *Arabidopsis*.

Example 11

5

The XVE System

In another series, the Lex A binding domain of the XVE system (pER12) is replaced by the zinc finger target DNA sequence from the chimera (see Figure 4) and the 8 LexA operator binding sites by 4 minimal binding zinc finger target DNA sequences. This new vector ZVE1 is tested by using either luciferase or GFP cloned
10 in the same vector in transgenic *Arabidopsis* plants.

The use of zinc finger polypeptides, chimeric proteins and the promoter systems described herein offers the advantages of great specificity and avoids non-specific interaction with the host genome.

All publications mentioned in the above specification are herein incorporated
15 by reference. Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such
20 specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

CLAIMS

1. A method of regulating transcription in a plant cell from a DNA sequence comprising a target DNA operably linked to a coding sequence, which method comprises introducing an engineered zinc finger polypeptide into said plant cell which polypeptide binds to the target DNA and modulates transcription of the coding sequence.
2. The method according to claim 1 wherein the target DNA is part of an endogenous genomic sequence.
3. The method according to claim 1 wherein the target DNA and coding sequence are heterologous to the cell.
4. The method according to any one of the preceding claims wherein the zinc finger polypeptide is fused to a biological effector domain.
5. The method according to claim 4 wherein the zinc finger polypeptide is fused to a transcriptional activator domain.
6. The method according to claim 4 wherein the zinc finger polypeptide is fused to a transcriptional repressor domain.
7. A plant host cell comprising a polynucleotide encoding an engineered zinc finger polypeptide and a target DNA sequence to which the zinc finger polypeptide binds.
8. A transgenic plant comprising a polynucleotide encoding an engineered zinc finger polypeptide and a target DNA sequence to which the zinc finger polypeptide binds.
9. A method according to any one of claim 1 to 6 wherein the plant cell is part of a plant and the target sequence is part of a regulatory sequence to which the nucleotide sequence of interest is operably linked.

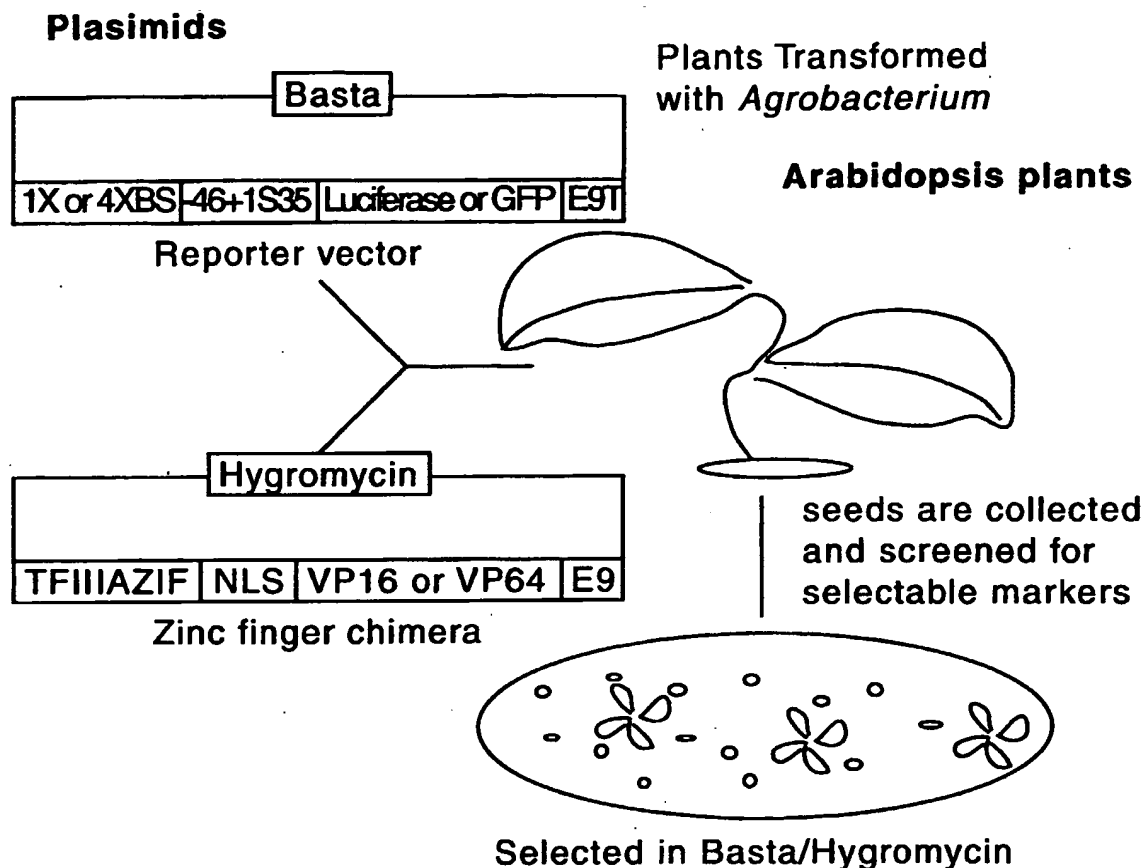
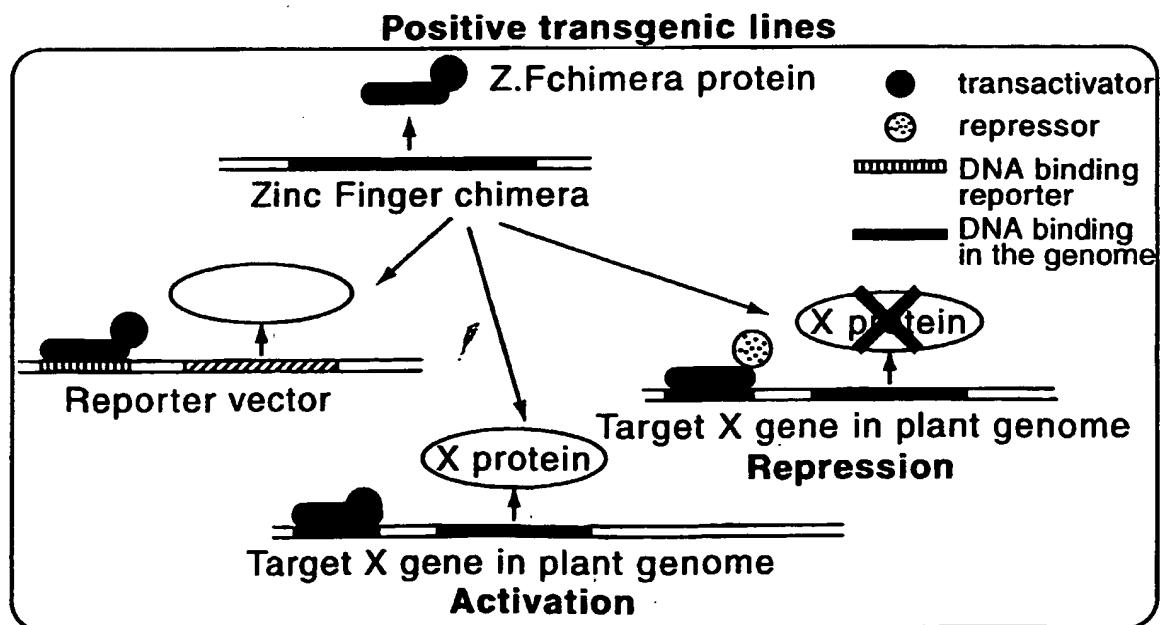
FIG. 1**FIG. 2**

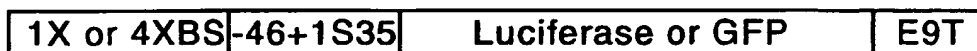
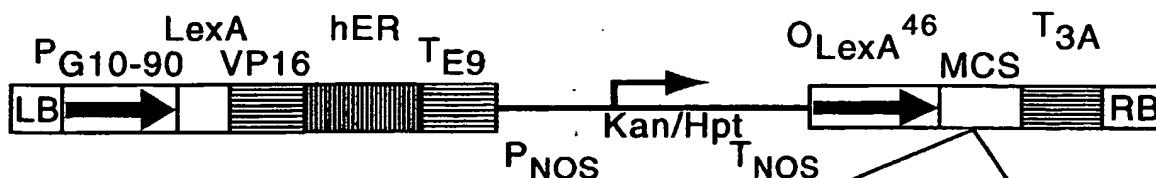
FIG. 3**Zinc finger chimera****Reporter construct****First series pBA002****Second Series pER8**

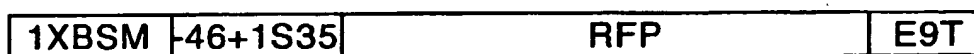
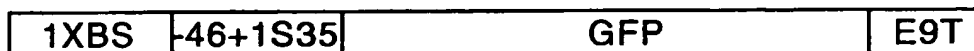
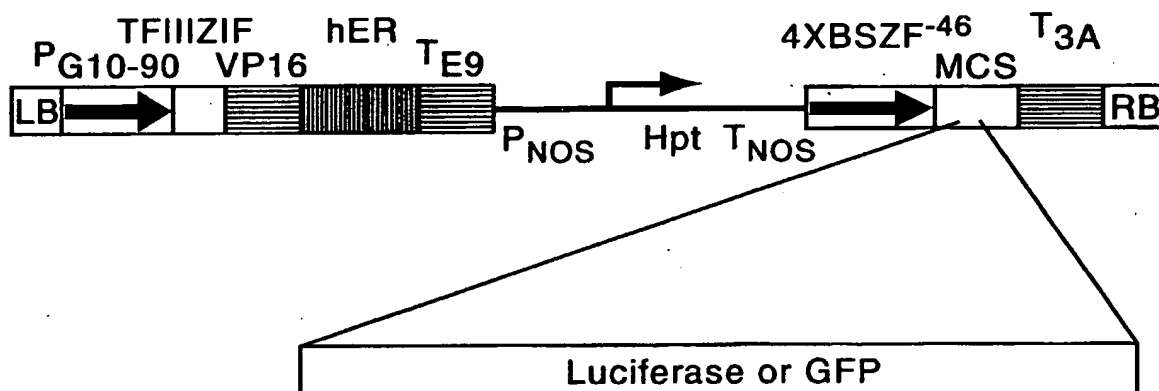
FIG. 4**Reporter construct of second series****Third series pZVE1**

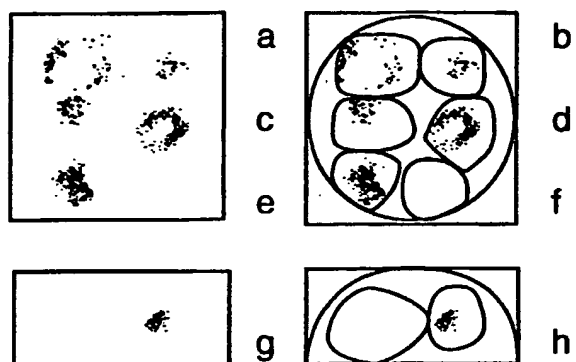
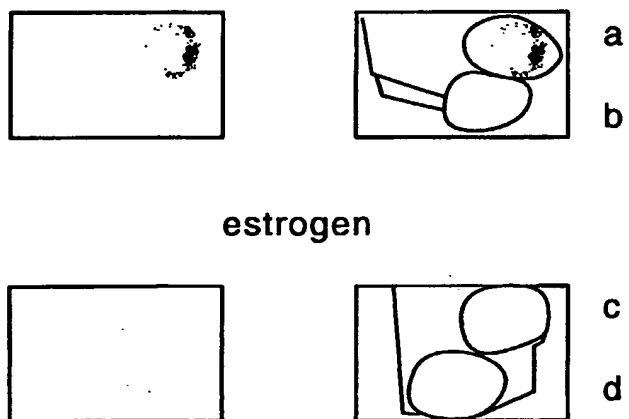
FIG. 5**FIG. 6**

FIG. 7

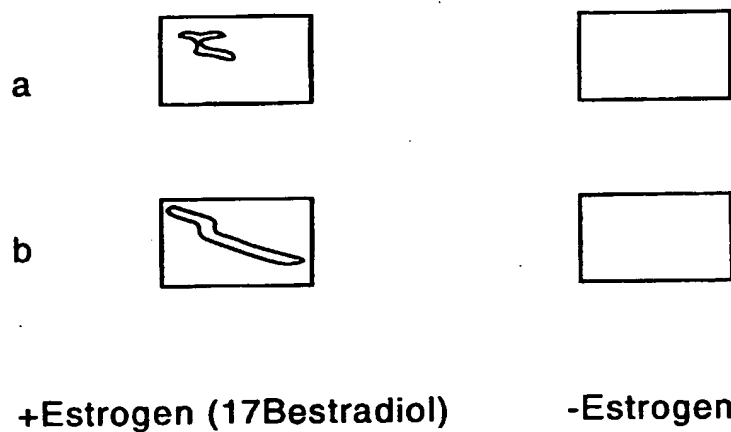


FIG. 8

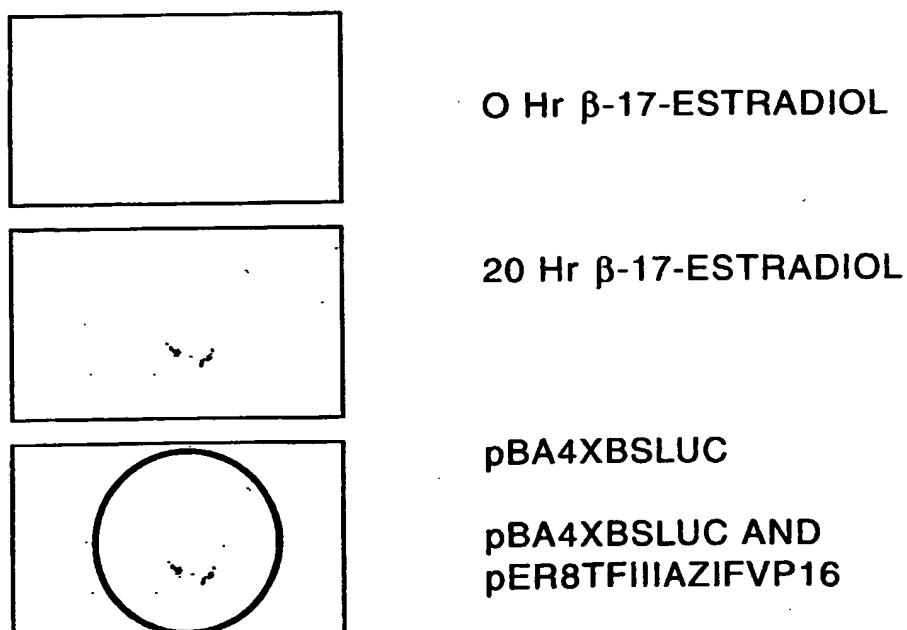


FIG. 9

	-46+S35	GFP	E9T
--	---------	-----	-----

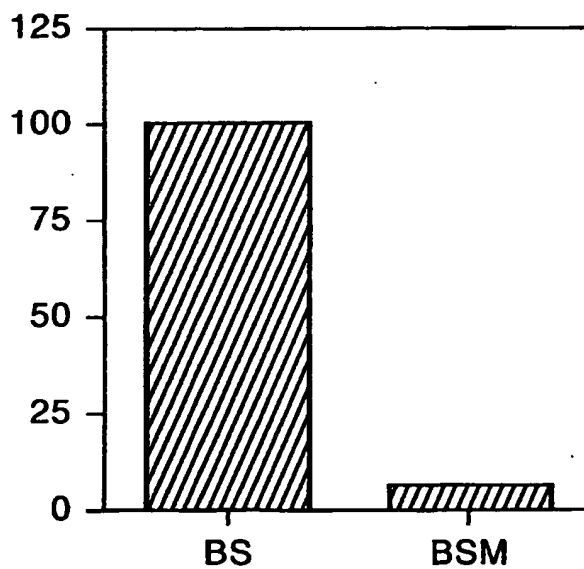
1XBS

GGTCTCCCATCAGGTACACGCCCACGCA

	-46+S35	RFP	E9T
--	---------	-----	-----

1XBSM (single mutation)

GGTCTCCCATCAGGTACACGCGCACGCA

%
positive cell %
positive cell

Constructs

Figure 10

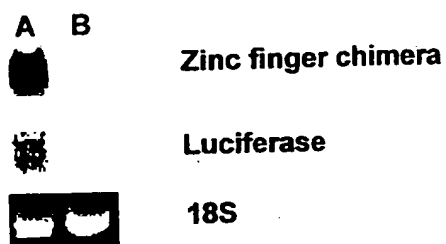
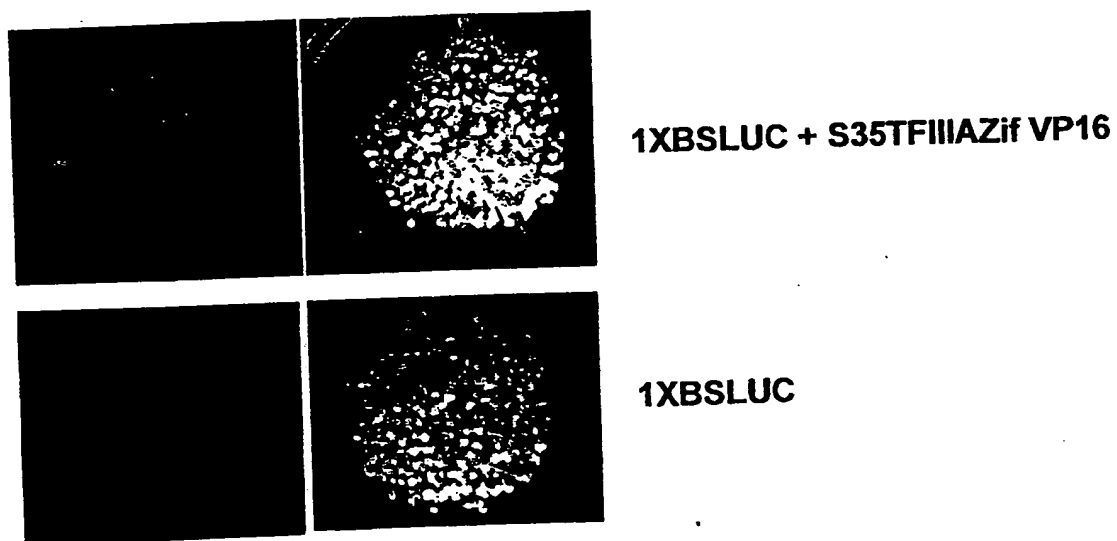


Figure 11



SEQUENCE LISTING

<110> Yen Choo, et al.

<120> Regulated Gene Expression in Plants

<130> 674538-2001.1WO

<160> 21

<170> PatentIn version 3.0

<210> 1

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> (1)..(14)

<223> Plant translational initiation context sequence

<400> 1

aaggagatat aacaatg

17

<210> 2

<211> 10

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> (1)..(7)

<223> plant translational initiation context sequence

<400> 2

gtcgaccatg

10

<210> 3

<211> 60

<212> DNA

<213> Artificial Sequence

<220>
<221> misc_feature
<222> (1)..(60)
<223> oligonucleotide

<400> 3
ctcctgcagt tggacctgtg ccatggccgg ctgggccgca tagaatggaa caactaaagc 60

<210> 4
<211> 995
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (15)..(17)
<223> translational initiating ATG

<220>
<221> misc_feature
<222> (16)..(416)
<223> Fingers 1 to 4 of TFIIIA

<220>
<221> misc_feature
<222> (308)..(416)
<223> spacer

<220>
<221> misc_feature
<222> (417)..(689)
<223> three fingers of zinc fingers protein Zif268

<220>
<221> misc_feature
<222> (701)..(722)
<223> Nuclear Localization Signal

<220>
<221> misc_feature

<222> (957) .. (986)

<223> c-myc tag

<400> 4

tctagagcgc cgccatggga gagaaggcgc tgccggtggt gtataagcgg tacatctgct	60
ctttcgccga ctgcggcgct gcttataaca agaactggaa actgcaggcg catctgtgca	120
aacacacagg agagaaacca tttccatgta aggaagaagg atgtgagaaa ggctttacct	180
cgcttcatca cttaaccgc cactcactca ctcatactgg cgagaaaaac ttcacatgtg	240
actcggatgg atgtgacttg agatttacta caaaggcaaa catgaagaag cactttaaca	300
gattccataa catcaagatc tgcgtctatg tgtgccattt tgagaactgt ggcaaagcat	360
tcaagaaaca caatcaatta aagggttcac agttcagtc cacacagcag ctgccgtatg	420
cttgccctgt cgagtcctgc gatcgccgct tttctcgctc ggatgagctt acccgccata	480
tccgcatcca cacaggccag aagcccttcc agtgtcgaat ctgcatgcgt aacttcagtc	540
gtagtgacca ccttaccacc cacatccgca cccacacagg cgagaagcct tttgcctgtg	600
acatttgtgg gaggaagttt gccaggagtg atgaacgcaa gaggcatacc aaaatccatt	660
taagacagaa ggacgcggcc gcactcgagc ggaattccgg cccaaaaaag aagagaaagg	720
tcgccccccc gaccgatgtc agcctggggg acgagctcca cttagacggc gaggacgtgg	780
cgatggcgca tgccgacgcg ctagacgatt tcgatctgga catgttgggg gacggggatt	840
ccccggggcc gggatttacc cccacgact ccgcccccta cggcgctctg gatacggccg	900
acttcgagtt tgagcagatg tttaccgatg cccttggaat tgacgagtac ggtggggaac	960
aaaaacttat ttctgaagaa gatctgtaag gatcc	995

<210> 5

<211> 947

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> (723)..(908)

<223> transactivation domain of VP64, other features except c-myc tag (listed below) same as SEQ ID NO:

<220>

<221> misc_feature

<222> (909)..(938)

<223> c-myc tag, other features except transactivation domain VP64 (listed above) same as SEQ ID NO:

<400> 5

tctagagcgc cgccatggga gagaaggcgc tgccggtggt gtataagcgg tacatctgct	60
ctttcgccga ctgcggcgct gcttataaca agaactggaa actgcaggcg catctgtgca	120
aacacacagg agagaaacca tttccatgta aggaagaagg atgtgagaaa ggctttacct	180
cgcttcatca cttaaccgcg cactcactca ctcatctgg cgagaaaaac ttcacatgtg	240
actcggatgg atgtgacttg agatttacta caaaggcaaa catgaagaag cactttaaca	300
gattccataa catcaagatc tgcgtctatg tgtgccattt tgagaactgt ggcaaagcat	360
tcaagaaaca caatcaatta aaggttcatc agttcagtca cacacagcag ctgccgatg	420
cttgccctgt cgagtcctgc gatcgccgct tttctcgctc ggatgagctt acccgccata	480
tccgcatcca cacaggccag aagcccttcc agtgtcgaat ctgcatgctg aacttcagtc	540
gtagtgacca ccttaccacc cacatccgca cccacacagg cgagaagcct tttgcctgtg	600
acatttgtgg gaggaagttt gccaggagtg atgaacgcaa gaggcatacc aaaatccatt	660
taagacagaa ggacgcggcc gcactcgagc ggaattccgg cccaaaaaag aagagaaagg	720
tcgaacttca gctgacttcg gatgcattag atgactttga cttagatatg ctaggatctg	780
acgcgctaga cgatttcgat ctggacatgt tgggcagcga tgctctagac gatttcgatt	840
tagatatgct tggctcggat gccctggatg acttcgacct cgacatgctg tcaagtcagc	900
tgagccagga acaaaaactt atttctgaag aagatctgta aggatcc	947

<210> 6

<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (1)..(14)
<223> plant translational initiation context sequence

<400> 6
aaggagatat aaca

14

<210> 7
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<221> protein_bind
<222> (1)..(29)
<223> target DNA sequence

<400> 7
tgcgtgggcg tgtacctgga tgggagacc

29

<210> 8
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (1)..(35)
<223> forward primer

<400> 8
ccacgcgtcc atgggagaga aggcgctgcc ggtgg

35

<210> 9
<211> 44
<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> (1)..(44)

<223> reverse primer

<400> 9

ccactagtc ttacagatct tcttcagaaa taagtttttg ttcc

44

<210> 10

<211> 148

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> (1)..(148)

<223> Sense strand primer

<400> 10

cctctagatc ggtctcccat ccaggtagac gccacgcaa gtcggtctcc catccaggta 60

cacgcccacg caagtcgggc tcccatccag gtacacgccc acgcaagtcg gtctcccatc 120

caggtagacg cccacgcaag aagcttcc 148

<210> 11

<211> 148

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> (1)..(148)

<223> antisense strand primer

<400> 11

ggaagcttct tgcgtgggcg tgtacctgga tgggagaccg acttgcgtag gcgtgtacct 60

ggatgggaga ccgacttgcg tgggcgtgta cctggatggg agaccgactt gcgtgggcgt 120

gtacctggat gggagaccga tctagagg

148

<210> 12
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (1)..(45)
<223> forward primer

<400> 12
ccagatctgg tctcccatcc aggtacacgc ccacgcaaga tctcc

45

<210> 13
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (1)..(46)
<223> reverse primer

<400> 13
ggagatcttg cgtgggcgtg tacctggatg ggagaccaga tctcgg

46

<210> 14
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (1)..(34)
<223> forward primer

<400> 14
ccccatgggtg agcaagggcg aggagctgtt cacc

34

<210> 15
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (1)..(35)
<223> reverse primer

<400> 15
ccgaattctt acttgtagag ctggtccatg ccgag

35

<210> 16
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (1)..(28)
<223> forward primer

<400> 16
ccctcgagcg gggtaccgag ggccccgg

28

<210> 17
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (1)..(30)
<223> reverse primer

<400> 17
cagttggaat tctagagtcg cggccgctac

30

<210> 18

<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (1)..(38)
<223> forward primer

<400> 18
ccgctcgagg ccccccgac cgatgtcagc ctggggga

38

<210> 19
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (1)..(38)
<223> reverse primer

<400> 19
ccgctcgagt attaatttga gaatgaacaa aaaggacc

38

<210> 20
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (1)..(38)
<223> forward primer

<400> 20
gccattaatc ggaatgggag agaaggcgct gccggtgg

38

<210> 21
<211> 32
<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> (1)..(32)

<223> reverse primer

<400> 21

gcctattaat ttgagaatga acaaaaagga cc

32